

```
!!SEQUENCE_LIST 1.0
!!FINDPATTERNS on PIR: * allowing 0 mismatches
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C,T,S)(R,K,H

PIR1:CCOF2T  ck: 2620  len: 92  finds: 1  | cytochrome c2 - Rhodocyclus re
PIR1:BDHCB  ck: 7216  len: 277  finds: 1  | carbonyl reductase (NADPH) (EC
PIR1:BDALAE ck: 2829  len: 246  finds: 1  | acetoacetyl-CoA reductase (EC
PIR1:S05397  ck: 3104  len: 272  finds: 1  | granaicin polyketide ketoredu
PIR1:S25079  ck: 4473  len: 261  finds: 1  | monensin polyketide ketoreduct
PIR1:BA2147  ck: 3492  len: 244  finds: 1  | 3-oxoacyl-[acyl-carrier-prote
PIR1:DB9930  ck: 2171  len: 238  finds: 1  | probable 3-oxoacyl-[acyl-carri
PIR1:RDECEP  ck: 6342  len: 334  finds: 1  | N-acetyl-gamma-glutamyl-phosph
PIR1:DTECC  ck: 4302  len: 311  finds: 1  | aspartate carboxyltransferase
PIR1:OWEBAC  ck: 4073  len: 311  finds: 1  | aspartate carboxyltransferase
PIR1:OWSEAC  ck: 6543  len: 306  finds: 1  | aspartate carboxyltransferase
PIR1:OWZP  ck: 9240  len: 327  finds: 1  | ornithine carboxyltransferase
PIR1:AG9962  ck: 158  len: 299  finds: 1  | phosphate butyryltransferase
PIR1:XNEBHC  ck: 1849  len: 359  finds: 1  | histidinol-phosphate transamin
PIR1:S29090  ck: 5401  len: 367  finds: 1  | dual specificity phosphoprotei
PIR1:S24411  ck: 5871  len: 367  finds: 1  | dual specificity phosphoprotei
PIR1:PRMSCL  ck: 92  len: 247  finds: 1  | cytotoxic T-lymphocyte protein
PIR1:OKSE  ck: 107  len: 376  finds: 1  | beta-lactamase (EC 3.5.2.6) pr
PIR1:DS3312  ck: 6718  len: 133  finds: 1  | cytidine deaminase (EC 3.5.4.3
PIR1:JE0022  ck: 9048  len: 136  finds: 1  | cytidine/deoxycytidine deamin
PIR1:F69500  ck: 4460  len: 219  finds: 1  | riboflavin-specific deaminase
PIR1:I3612  ck: 2178  len: 142  finds: 1  | H+-transporting ATP synthase
PIR1:RKIOS2  ck: 4535  len: 180  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKIOS1  ck: 5240  len: 181  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKPOS1  ck: 4699  len: 181  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:RKPOS8  ck: 2967  len: 180  finds: 1  | ribulose-bisphosphate carboxyl
PIR1:F69398  ck: 4436  len: 277  finds: 1  | dihydroxynaphthoic acid syntha
PIR1:F71428  ck: 9250  len: 244  finds: 1  | hypothetical protein - Arabid
PIR1:A48681  ck: 300  len: 376  finds: 1  | placental thrombin inhibitor
PIR1:WIWTA  ck: 4689  len: 123  finds: 1  | alpha-amyase inhibitor CIIT
PIR1:WIWTA5  ck: 4073  len: 124  finds: 1  | alpha-amyase inhibitor 0.53
PIR1:WIWTA1  ck: 3320  len: 124  finds: 1  | alpha-amyase inhibitor 0.19
PIR1:TVHUP1  ck: 4677  len: 380  finds: 1  | transforming protein fos - hum

PIR1:TVCHRS  ck: 7007  len: 367  finds: 1  | transforming protein fos -
PIR1:TVFEF4  ck: 9324  len: 322  finds: 1  | transforming protein fos -
PIR1:S00755  ck: 1536  len: 350  finds: 1  | pleckstrin - human
PIR1:HLHUG  ck: 908  len: 216  finds: 1  | class II histocompatibility
PIR1:RLJLV8  ck: 6002  len: 132  finds: 1  | ribosomal protein S8 - live
PIR1:S26396  ck: 6968  len: 151  finds: 1  | ribosomal protein S13-e, cy
PIR1:A48103  ck: 718  len: 256  finds: 1  | ribosomal protein L21 precu
PIR1:QJ1298  ck: 409  len: 340  finds: 1  | annexin II type 2 - African
PIR1:OOFF  ck: 4358  len: 373  finds: 1  | rhodopsin - fruit fly (Dros
PIR1:I55604  ck: 4750  len: 411  finds: 1  | platelet glycoprotein Ib be
PIR1:OREBOB  ck: 4920  len: 306  finds: 1  | oligopeptide transport syst
PIR1:S67566  ck: 8415  len: 417  finds: 1  | probable membrane protein Y
PIR1:S30896  ck: 7059  len: 297  finds: 1  | virulence-associated protei
PIR1:S06670  ck: 7032  len: 297  finds: 1  | virulence-associated protei
PIR1:IMECNA  ck: 6528  len: 131  finds: 1  | colicin N immunity protein
PIR1:ZMECD  ck: 8403  len: 139  finds: 1  | proteinase unmd (EC 3.4.21.
PIR1:HE4563  ck: 9038  len: 343  finds: 1  | flagellar motor switch prot
PIR1:BVECU  ck: 4271  len: 130  finds: 1  | unci protein - Escherichia
PIR1:BBAG55  ck: 173  len: 257  finds: 1  | virB8 protein precursor - A
PIR1:BBAG46  ck: 2646  len: 230  finds: 1  | virB8 protein - Agrobacteri
PIR1:BBAG58  ck: 6263  len: 237  finds: 1  | virB8 protein - Agrobacteri
PIR1:WZWL31  ck: 9656  len: 372  finds: 1  | E2 protein - human papillom
PIR1:WZWL51  ck: 1816  len: 358  finds: 1  | E2 protein - human papillom
PIR1:WZWL18  ck: 5567  len: 365  finds: 1  | E2 protein - human papillom
PIR1:SI5617  ck: 734  len: 391  finds: 1  | E2 protein - human papillom
PIR1:QOBENS  ck: 7071  len: 366  finds: 1  | integral membrane protein -
PIR1:WNN222  ck: 9828  len: 194  finds: 1  | envelope-associated 22k pro
PIR1:WNN2BA  ck: 6337  len: 186  finds: 1  | matrix glycoprotein M2 - bo
PIR1:WNVYP2  ck: 3415  len: 109  finds: 1  | 12k protein - potato virus
PIR1:VCBWC  ck: 3595  len: 237  finds: 1  | coat protein - tobacco stre
PIR1:VCVETC  ck: 1818  len: 351  finds: 1  | coat protein - turnip crink
PIR1:VHBP6  ck: 6252  len: 149  finds: 1  | major nucleocapsid protein
PIR1:ZGBG4  ck: 4484  len: 177  finds: 1  | gene G protein - phage G4
PIR1:PDXYC  ck: 6596  len: 296  finds: 1  | REP2 protein - yeast (Sacch
PIR1:BE4166  ck: 1216  len: 214  finds: 1  | hypothetical protein H1084
PIR1:E70044  ck: 4650  len: 310  finds: 1  | conserved hypothetical prot
```

PIR1:G69000	ck: 7972	len: 409	finds: 1	molybdenum cofactor biosynthes	PIR2:T09997	ck: 1856	len: 385	finds: 1	lipoxigenase (EC 1.13.11.12
PIR1:E70191	ck: 8273	len: 337	finds: 1	conserved hypothetical protein	PIR2:S43771	ck: 2848	len: 347	finds: 1	phosphatidylcholine desatur
PIR1:A71183	ck: 8100	len: 306	finds: 1	probable UDP-glucose 4-epimerase	PIR2:S74692	ck: 6210	len: 231	finds: 1	tRNA(mIG37)methyltransferase
PIR1:S70955	ck: 8920	len: 288	finds: 1	ctnf protein - Vbprlo cholerae	PIR2:A71119	ck: 9980	len: 308	finds: 1	probable aspartate carbamoy
PIR1:A70118	ck: 6242	len: 299	finds: 1	glycine betaine, L-proline ABC	PIR2:A70959	ck: 1547	len: 319	finds: 1	probable aspartate carbamoy
PIR2:S76375	ck: 6628	len: 126	finds: 1	plasmocyanin precursor - Syne	PIR2:C70815	ck: 556	len: 403	finds: 1	probable beta-ketoadipyl Co
PIR2:F70646	ck: 8411	len: 323	finds: 1	probable quinone oxidoreductas	PIR2:T12981	ck: 8558	len: 452	finds: 1	glucuronosyl transferase-11
PIR2:S01901	ck: 8887	len: 254	finds: 1	alcohol dehydrogenase (EC 1.1.	PIR2:T140986	ck: 7621	len: 323	finds: 1	streptomycin/spectinomycin
PIR2:S65732	ck: 4453	len: 132	finds: 1	carbonyl reductase (NADPH) (EC	PIR2:I65760	ck: 800	len: 285	finds: 1	alcohol sulfoxyltransferase (E
PIR2:SC5284	ck: 7185	len: 277	finds: 1	carbonyl reductase (NADPH) (EC	PIR2:S52265	ck: 6443	len: 367	finds: 1	dual specificity phosphatas
PIR2:SS4815	ck: 5190	len: 261	finds: 1	urdamycin polyketide ketoreduc	PIR2:T12078	ck: 3061	len: 231	finds: 1	ribonuclease (EC 3.1.-.-) -
PIR2:D55587	ck: 3920	len: 261	finds: 1	griseusin polyketide ketoreduc	PIR2:B71518	ck: 3917	len: 167	finds: 1	probable lipoprotein signal
PIR2:S69225	ck: 7076	len: 262	finds: 1	nogalamycin polyketide ketored	PIR2:S26429	ck: 9885	len: 178	finds: 1	probable dUTP pyrophosphata
PIR2:S35196	ck: 9329	len: 261	finds: 1	hypothetical protein 3 - Sacch	PIR2:S42592	ck: 4170	len: 131	finds: 1	urease (EC 3.5.1.5), tissue
PIR2:D64051	ck: 4159	len: 242	finds: 1	3-oxoacyl-[acyl-carrier-prote	PIR2:I51663	ck: 4321	len: 360	finds: 1	arginase 1 - African clawed
PIR2:D69648	ck: 3484	len: 254	finds: 1	2-keto-3-deoxygluconate oxidor	PIR2:I51664	ck: 4641	len: 360	finds: 1	arginase 2 - African clawed
PIR2:C48674	ck: 5293	len: 268	finds: 1	tropinone reductase homolog -	PIR2:I51665	ck: 4724	len: 360	finds: 1	arginase 3 - African clawed
PIR2:T12020	ck: 9922	len: 247	finds: 1	3-oxoacyl-acyl carrier protein	PIR2:J06069	ck: 8425	len: 140	finds: 1	blastididin-S deaminase (EC
PIR2:T12051	ck: 5984	len: 244	finds: 1	3-oxoacyl-[acyl-carrier-prote	PIR2:S37304	ck: 2135	len: 335	finds: 1	spal protein - Salmonella t
PIR2:S39654	ck: 5853	len: 255	finds: 1	hypothetical protein X - Pseud	PIR2:S01397	ck: 4392	len: 127	finds: 1	H+-transporting ATP synthas
PIR2:H69885	ck: 6790	len: 242	finds: 1	3-oxoacyl-acyl-carrier prote	PIR2:S22348	ck: 1686	len: 168	finds: 1	H+-transporting ATP synthas
PIR2:F69868	ck: 7200	len: 248	finds: 1	glucose 1-dehydrogenase homo	PIR2:D72121	ck: 2102	len: 209	finds: 1	ATP synthase chain D - Chla
PIR2:BS5850	ck: 1047	len: 267	finds: 1	hypothetical protein (hsdh 5'	PIR2:S31498	ck: 4433	len: 181	finds: 1	ribulose-bisphosphate carbo
PIR2:AS6275	ck: 101	len: 387	finds: 1	1,3-propanediol dehydrogenase	PIR2:A71295	ck: 255	len: 357	finds: 1	probable alanine racemase (
PIR2:S47829	ck: 4443	len: 339	finds: 1	glycerol-3-phosphate dehydroge	PIR2:G70553	ck: 8877	len: 268	finds: 1	probable echA10 protein - M
PIR2:A44132	ck: 6742	len: 338	finds: 1	D-aspartate oxidase (EC 1.4.3.	PIR2:H69457	ck: 4102	len: 322	finds: 1	ornithine cyclodeaminase (a
PIR2:SC5438	ck: 7167	len: 341	finds: 1	D-aspartate oxidase (EC 1.4.3.	PIR2:S75089	ck: 8968	len: 338	finds: 1	UDP-glucose 4-epimerase - S
PIR2:SC5439	ck: 934	len: 282	finds: 1	D-aspartate oxidase (EC 1.4.3.	PIR2:H69105	ck: 3451	len: 336	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:S25422	ck: 6471	len: 72	finds: 1	NADH dehydrogenase (ubiquinone	PIR2:D66290	ck: 128	len: 332	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:S08622	ck: 675	len: 307	finds: 1	hydrogennase (EC 1.18.99.1) 3	PIR2:S47045	ck: 5086	len: 346	finds: 1	dUDP-glucose 4,6-dehydratas
PIR2:T11338	ck: 2220	len: 347	finds: 1	NADH dehydrogenase (ubiquinone	PIR2:E70566	ck: 5640	len: 331	finds: 1	probable DTDP-GLUCOSE 4 - M
PIR2:T11159	ck: 979	len: 114	finds: 1	NADH dehydrogenase 3 - hardbac	PIR2:G70415	ck: 2675	len: 321	finds: 1	nucleotide sugar epimerase
PIR2:S42244	ck: 1655	len: 214	finds: 1	NADH dehydrogenase (ubiquinone	PIR2:E69750	ck: 3388	len: 249	finds: 1	glucosamine-6-phosphate iso
PIR2:F64959	ck: 2909	len: 184	finds: 1	NADH dehydrogenase (ubiquinone	PIR2:D69094	ck: 5550	len: 248	finds: 1	phosphoribosylaminoimidazol
PIR2:S52870	ck: 6545	len: 167	finds: 1	NADH dehydrogenase (ubiquinone	PIR2:A19940	ck: 1681	len: 40	finds: 1	antithrombin III - baboon (
PIR2:S78138	ck: 8115	len: 267	finds: 1	cytochrome-c oxidase (EC 1.9.3	PIR2:J0X206	ck: 7697	len: 207	finds: 1	chymotrypsin inhibitor (Kun
PIR2:S07557	ck: 7659	len: 268	finds: 1	cytochrome-c oxidase (EC 1.9.3					

PIR2:JS0650	ck: 8167	len: 207	finds: 1	i chymotrypsin inhibitor (Kunitz
PIR2:JX0246	ck: 1378	len: 181	finds: 1	i serine proteinase inhibitor A
PIR2:JCS447	ck: 4632	len: 210	finds: 1	i serine proteinase inhibitor A
PIR2:T06517	ck: 7537	len: 151	finds: 1	i alpha-amylase inhibitor 1mal F
PIR2:SI6920	ck: 2886	len: 121	finds: 1	i alpha-amylase inhibitor - whea
PIR2:AE0195	ck: 2543	len: 278	finds: 1	i transforming protein dbl - mou
PIR2:AA0316	ck: 7358	len: 415	finds: 1	i cyclin B - fission yeast (Schl
PIR2:AS7234	ck: 2377	len: 348	finds: 1	i lin-44 protein precursor - Cae
PIR2:JCA152	ck: 4630	len: 354	finds: 1	i Wnt-11 protein precursor - chi
PIR2:SS4378	ck: 1689	len: 354	finds: 1	i wnt-11 protein - mouse
PIR2:JH0688	ck: 1620	len: 398	finds: 1	i bone morphogenetic protein 211
PIR2:JH0687	ck: 9563	len: 398	finds: 1	i bone morphogenetic protein 21
PIR2:SS21299	ck: 8703	len: 81	finds: 1	i protein DAF8 - eastern green m
PIR2:AS37045	ck: 6756	len: 189	finds: 1	i thermostable direct hemolysin
PIR2:AA8942	ck: 5895	len: 189	finds: 1	i thermostable direct hemolysin
PIR2:JQ0472	ck: 2831	len: 135	finds: 1	i T-cell receptor beta chain (BT
PIR2:JQ0473	ck: 3911	len: 136	finds: 1	i T-cell receptor beta chain (BT
PIR2:BS5250	ck: 6952	len: 353	finds: 1	i class I histocompatibility ant
PIR2:IS0609	ck: 8972	len: 207	finds: 1	i T-cell surface glycoprotein CD
PIR2:IS39516	ck: 7310	len: 257	finds: 1	i ribosomal protein S3 - Achroma
PIR2:SS77489	ck: 8926	len: 133	finds: 1	i ribosomal protein S8 - Synech
PIR2:T01947	ck: 4063	len: 252	finds: 1	i hypothetical protein FlloA.1 -
PIR2:AS9234	ck: 9791	len: 371	finds: 1	i opsin - bluebottle fly (Calli
PIR2:SS4091	ck: 7181	len: 374	finds: 1	i opsin rh1 - fruit fly (Drosoph
PIR2:JCA304	ck: 2863	len: 355	finds: 1	i orphan G-protein-coupled recep
PIR2:IS58186	ck: 4478	len: 354	finds: 1	i hypothetical G-protein coupled
PIR2:AS9667	ck: 8980	len: 60	finds: 1	i pulmonary surfactant protein B
PIR2:SS2494	ck: 429	len: 277	finds: 1	i rRNA N-glycosidase (EC 3.2.2.2
PIR2:SS2319	ck: 9157	len: 278	finds: 1	i beta-luffin - smooth loofah
PIR2:JN0108	ck: 7681	len: 250	finds: 1	i luffin-b - smooth loofah
PIR2:SS65052	ck: 5285	len: 214	finds: 1	i pistril-specific protein sts14
PIR2:JCA822	ck: 8148	len: 77	finds: 1	i acyl carrier protein - Bacillus
PIR2:BS65003	ck: 7651	len: 228	finds: 1	i histidine transport system per
PIR2:BS6263	ck: 5147	len: 306	finds: 1	i oligopeptide transport system
PIR2:C71163	ck: 4618	len: 304	finds: 1	i probable oligopeptide transport
PIR2:SS64728	ck: 6546	len: 482	finds: 1	i protein secretion protein xcp1
PIR2:SS2163	ck: 8969	len: 331	finds: 1	i sucrose specific repressor - B
PIR2:D70044	ck: 4428	len: 243	finds: 1	i transcription regulator Gnt
PIR2:SS48826	ck: 9647	len: 215	finds: 1	i NADH dehydrogenase (ubiquin
PIR2:SS75086	ck: 4153	len: 342	finds: 1	i iron-stress chlorophyll-b1n
PIR2:AS6434	ck: 3224	len: 182	finds: 1	i probable fibinrial protein-1
PIR2:D71484	ck: 1871	len: 179	finds: 1	i probable ribosome releasing
PIR2:SS74877	ck: 7310	len: 328	finds: 1	i phosphate transport system
PIR2:AA6259	ck: 9097	len: 438	finds: 1	i recombination protein reca
PIR2:JQ0661	ck: 6434	len: 424	finds: 1	i impb protein - Salmonella t
PIR2:F71946	ck: 9168	len: 343	finds: 1	i flagellar motor switch prot
PIR2:AS37763	ck: 2312	len: 147	finds: 1	i virD1 protein - Agrobacteri
PIR2:AS25063	ck: 4145	len: 147	finds: 1	i hypothetical virD1 protein
PIR2:SS06883	ck: 2761	len: 147	finds: 1	i virD1 protein - Agrobacteri
PIR2:AS36564	ck: 5828	len: 368	finds: 1	i E2 protein - human papillom
PIR2:SS36576	ck: 4491	len: 368	finds: 1	i E2 protein - human papillom
PIR2:SS36552	ck: 4819	len: 383	finds: 1	i E2 protein - human papillom
PIR2:SS36535	ck: 273	len: 376	finds: 1	i E2 protein - human papillom
PIR2:AA4215	ck: 5996	len: 303	finds: 1	i EUS1 protein - equine herpe
PIR2:F43674	ck: 501	len: 372	finds: 1	i US7 protein - human herpesv
PIR2:SS58346	ck: 2806	len: 256	finds: 1	i coat protein V1 - tomato ye
PIR2:Q1248	ck: 566	len: 106	finds: 1	i hypothetical 12K protein -
PIR2:AS1285	ck: 3148	len: 355	finds: 1	i coat protein - sweet potato
PIR2:ES71669	ck: 3620	len: 433	finds: 1	i preprotein translocase secy
PIR2:CE9632	ck: 4284	len: 258	finds: 1	i transcription repressor glic
PIR2:F69260	ck: 6950	len: 244	finds: 1	i nitrate ABC transporter, pe
PIR2:A70001	ck: 6987	len: 253	finds: 1	i ABC transporter (ATP-bindin
PIR2:A70039	ck: 1714	len: 301	finds: 1	i ABC transporter (ATP-bindin
PIR2:SS69260	ck: 9269	len: 243	finds: 1	i nitrate ABC transporter, AT
PIR2:SS6325	ck: 1039	len: 252	finds: 1	i hisp-like nucleotide bindin
PIR2:F73545	ck: 2944	len: 450	finds: 1	i ATP-dependent RNA helicase
PIR2:SS69240	ck: 2958	len: 482	finds: 1	i probable RNA helicase CAS/6
PIR2:AS8990	ck: 3980	len: 272	finds: 1	i T-cell-specific transcript1
PIR2:JH0402	ck: 9457	len: 279	finds: 1	i T-cell-specific transcript1
PIR2:SS0068	ck: 241	len: 393	finds: 1	i nonhistone chromosomal prot
PIR2:AS8095	ck: 5007	len: 269	finds: 1	i T-cell-specific transcript1
PIR2:BS8095	ck: 3201	len: 269	finds: 1	i T-cell-specific transcript1
PIR2:D38095	ck: 7876	len: 250	finds: 1	i T-cell-specific transcript1

PIR2:CG8095	ck: 409	len: 268	finds: 1	T-cell-specific transcription	PIR2:D64151	ck: 1410	len: 430	finds: 1	hypothetical protein HI0404
PIR2:JC6179	ck: 83	len: 393	finds: 1	dorsal switch protein 1 - fruit	PIR2:S75967	ck: 3587	len: 271	finds: 1	hypothetical protein - Syne
PIR2:SS1796	ck: 5713	len: 366	finds: 1	T-cell-specific transcription	PIR2:D69014	ck: 6260	len: 222	finds: 1	conserved hypothetical prot
PIR2:JN0767	ck: 3795	len: 270	finds: 1	homeobox protein HEX - human	PIR2:CG5012	ck: 4986	len: 285	finds: 1	hypothetical protein b2382
PIR2:I46089	ck: 4988	len: 371	finds: 1	thyroid transcription factor-1	PIR2:A70606	ck: 144	len: 358	finds: 1	probable electron transfer
PIR2:SS0230	ck: 8616	len: 271	finds: 1	homeotic protein Hex - mouse	PIR2:CG5112	ck: 597	len: 220	finds: 1	sigma cross-reacting protei
PIR2:SS3724	ck: 7622	len: 372	finds: 1	thyroid-specific enhancer-bin	PIR2:JC5603	ck: 1627	len: 270	finds: 1	Esi protein - zebra fish
PIR2:AS6451	ck: 5439	len: 371	finds: 1	thyroid-specific enhancer-bind	PIR2:T09924	ck: 1428	len: 293	finds: 1	cytidine deaminase (EC 3.5.
PIR2:SI2002	ck: 6608	len: 372	finds: 1	thyroid nuclear factor 1 - rat	PIR2:SS5551	ck: 2728	len: 257	finds: 1	cytokine-inducible protein
PIR2:IS9234	ck: 9010	len: 420	finds: 1	octamer binding transcription	PIR2:JC5626	ck: 1738	len: 198	finds: 1	STAT induced STAT inhibitor
PIR2:G02321	ck: 1141	len: 401	finds: 1	thyroid transcription factor 1	PIR2:JC5760	ck: 1381	len: 198	finds: 1	cytokine-inducible SR2 prot
PIR2:S78452	ck: 3745	len: 331	finds: 1	POU-domain protein rdc-1 - hum	PIR2:A41900	ck: 4304	len: 299	finds: 1	cyn operon regulatory prote
PIR2:S76329	ck: 4655	len: 416	finds: 1	probable phosphoesterase (EC 3	PIR2:F70721	ck: 7074	len: 164	finds: 1	hypothetical protein RV1829
PIR2:F71508	ck: 6469	len: 244	finds: 1	probable phosphoesterase (EC 3	PIR2:SS6618	ck: 3873	len: 173	finds: 1	yjyx protein - Escherichia
PIR2:S35782	ck: 5295	len: 467	finds: 1	serine/threonine protein kinase	PIR2:S45255	ck: 4167	len: 54	finds: 1	yjyx protein homolog - Ente
PIR2:CG9230	ck: 3724	len: 257	finds: 1	conserved hypothetical protein	PIR2:A70738	ck: 8373	len: 158	finds: 1	probable rimi protein - Myc
PIR2:AA0369	ck: 4295	len: 302	finds: 1	trans-acting transcription reg	PIR2:F72316	ck: 5138	len: 309	finds: 1	hypothetical protein TM0917
PIR2:AA3960	ck: 1522	len: 306	finds: 1	cytochrome caa3 oxidase ctaa-	PIR2:B70451	ck: 4885	len: 311	finds: 1	prephenate dehydrogenase -
PIR2:AA5335	ck: 368	len: 297	finds: 1	cytochrome-c oxidase assembly	PIR2:G39845	ck: 2658	len: 256	finds: 1	dihydroorotate dehydrogenas
PIR2:G02514	ck: 2327	len: 326	finds: 1	P2 purinoceptor - human	PIR2:JO1382	ck: 6785	len: 303	finds: 1	hypothetical 3A4 protein -
PIR2:JC4800	ck: 7545	len: 328	finds: 1	P26 receptor - human	PIR2:SS6961	ck: 8224	len: 196	finds: 1	probable membrane protein y
PIR2:FF4471	ck: 3696	len: 415	finds: 1	hypothetical protein homolog M	PIR2:JO1355	ck: 9413	len: 134	finds: 1	v0 protein - Muscathus str
PIR2:AA7201	ck: 1219	len: 120	finds: 1	bombesinlike peptide - African	PIR2:JO0150	ck: 6657	len: 122	finds: 1	hypothetical 13k protein -
PIR2:BS5139	ck: 1678	len: 197	finds: 1	hypothetical protein B3434 - E	PIR2:B71217	ck: 8463	len: 108	finds: 1	hypothetical protein PH2002
PIR2:SA48737	ck: 6122	len: 423	finds: 1	kynurenine aminotransferase -	PIR2:T05820	ck: 4391	len: 217	finds: 1	hypothetical protein TSK18.
PIR2:JO2268	ck: 6645	len: 364	finds: 1	O-methyltransferase (EC 2.1.1.	PIR2:S77462	ck: 3256	len: 275	finds: 1	hypothetical protein s11107
PIR2:G70165	ck: 3950	len: 315	finds: 1	aldose reductase homolog - Lym	PIR2:SS3572	ck: 4181	len: 158	finds: 1	probable membrane protein y
PIR2:G69076	ck: 4836	len: 251	finds: 1	molybdopterin biosynthesis pro	PIR2:GA5335	ck: 1335	len: 146	finds: 1	ORF1 protein - Autographa c
PIR2:SS6490	ck: 2843	len: 332	finds: 1	hypothetical transcription reg	PIR2:SS2975	ck: 4975	len: 383	finds: 1	gene BCR2 protein - human
PIR2:HG9823	ck: 1929	len: 302	finds: 1	conserved hypothetical protein	PIR2:SO2333	ck: 5478	len: 57	finds: 1	probable membrane antigen C
PIR2:AA4223	ck: 590	len: 308	finds: 1	probable pseudouridine synthas	PIR2:SO3243	ck: 3042	len: 115	finds: 1	hypothetical protein B-115
PIR2:S73869	ck: 6533	len: 309	finds: 1	hypothetical protein ycec - My	PIR2:SS21504	ck: 7782	len: 148	finds: 1	hypothetical protein (gene
PIR2:BB4026	ck: 2974	len: 164	finds: 1	hypothetical protein H11338 -	PIR2:SS5282	ck: 7758	len: 317	finds: 1	eaa protein - phage P22
PIR2:BB5007	ck: 9360	len: 161	finds: 1	hypothetical protein b2340 - E	PIR2:SI7834	ck: 1383	len: 36	finds: 1	acetyl-CoA carboxylase - ph
PIR2:CG4882	ck: 51	len: 299	finds: 1	probable transcription regulat	PIR2:AS6663	ck: 6051	len: 299	finds: 1	capsid protein VPI - foot-a
PIR2:ET2058	ck: 4148	len: 202	finds: 1	probable phosphatase/kinase -	PIR2:HT2245	ck: 6860	len: 236	finds: 1	2,3,4,5-tetrahydropyridine-
PIR2:BV1509	ck: 4232	len: 202	finds: 1	probable phosphatase/kinase -					

PIR2:E72316	ck: 6581	len: 410	finds: 1	! hypothetical protein - Thermot
PIR2:BA0609	ck: 8809	len: 236	finds: 1	! probable lipoprotein MlpA - My
PIR2:JC6019	ck: 2747	len: 290	finds: 1	! response-regulator protein Fir
PIR2:CT0198	ck: 1619	len: 440	finds: 1	! conserved hypothetical protein
PIR2:E70167	ck: 2658	len: 192	finds: 1	! hypothetical protein BB0542 -
PIR2:BT0174	ck: 8103	len: 208	finds: 1	! hypothetical protein BB0595 -
PIR2:DT1926	ck: 2840	len: 142	finds: 1	! cag island protein - Helicobac
PIR2:F71925	ck: 6963	len: 114	finds: 1	! cag island protein - Helicobac
PIR2:HE4586	ck: 7347	len: 114	finds: 1	! cag pathogenicity island prote
PIR2:FE4587	ck: 2551	len: 142	finds: 1	! cag pathogenicity island prote
PIR2:AE4649	ck: 6457	len: 131	finds: 1	! hypothetical protein HP1033 -
PIR2:DE4710	ck: 2117	len: 115	finds: 1	! hypothetical protein HP1524 -
PIR2:ET1938	ck: 3806	len: 91	finds: 1	! hypothetical protein jhp0391 -
PIR2:AT1809	ck: 3881	len: 115	finds: 1	! hypothetical protein jhp1413 -
PIR2:CT1912	ck: 428	len: 307	finds: 1	! probable outer membrane protei
PIR2:PH0856	ck: 7474	len: 277	finds: 1	! mauf protein - Paracoccus deni
PIR2:IT9703	ck: 4597	len: 102	finds: 1	! tram protein - Agrobacterium t
PIR2:JC5568	ck: 7036	len: 118	finds: 1	! mxy protein - Rhizobium meli
PIR2:SZ7344	ck: 1548	len: 370	finds: 1	! hupK protein - Rhizobium legum
PIR2:SZ8677	ck: 6944	len: 243	finds: 1	! hypothetical protein 4 - Rhizoc
PIR2:SZ6139	ck: 4831	len: 187	finds: 1	! signalling protein ampd - Citr
PIR2:AA8901	ck: 4831	len: 187	finds: 1	! signalling protein ampd - Ente
PIR2:SA0867	ck: 3905	len: 248	finds: 1	! ferredoxin--NADP+ reductase (E
PIR2:AE4750	ck: 8141	len: 79	finds: 1	! hypothetical protein b0249 - E
PIR2:CE5059	ck: 1867	len: 425	finds: 1	! hypothetical protein b2775 - E
PIR2:BE5068	ck: 6883	len: 145	finds: 1	! hypothetical protein b2849 - E
PIR2:FE4963	ck: 7029	len: 359	finds: 1	! nicotinate-nucleotide--dimethy
PIR2:T00211	ck: 5937	len: 291	finds: 1	! Type II secretion pathway rela
PIR2:HE4726	ck: 98	len: 216	finds: 1	! yabP protein - Escherichia col
PIR2:AE4862	ck: 1943	len: 78	finds: 1	! ycgZ protein - Escherichia col
PIR2:AE4882	ck: 6021	len: 262	finds: 1	! ycjI protein - Escherichia col
PIR2:SZ0452	ck: 9681	len: 271	finds: 1	! hypothetical protein X - Klebs
PIR2:SO1838	ck: 7149	len: 220	finds: 1	! nifH protein - Klebsiella pneu
PIR2:SZ2619	ck: 7734	len: 336	finds: 1	! hypothetical protein - Salmon
PIR2:AS3302	ck: 5663	len: 289	finds: 1	! hemF 5'-region hypothetical pr
PIR2:SZ2906	ck: 4054	len: 135	finds: 1	! hypothetical protein 1 - Shiga
PIR2:TI4663	ck: 2204	len: 99	finds: 1	! histone H5 like protein - Yers
PIR2:TI4952	ck: 2530	len: 144	finds: 1	! hypothetical protein - Yers
PIR2:ST0883	ck: 2492	len: 227	finds: 1	! hypothetical protein 3 - Vi
PIR2:SZ54442	ck: 6509	len: 150	finds: 1	! membrane associated protein
PIR2:G64102	ck: 8973	len: 288	finds: 1	! phosphatidate cytidyltran
PIR2:HE64157	ck: 4	len: 311	finds: 1	! sufi protein homolog HI0733
PIR2:HT1692	ck: 1607	len: 405	finds: 1	! hypothetical protein RP358
PIR2:E72060	ck: 2665	len: 252	finds: 1	! ct470 hypothetical protein
PIR2:F72031	ck: 6017	len: 493	finds: 1	! ct578 hypothetical protein
PIR2:SE6492	ck: 5418	len: 270	finds: 1	! hypothetical protein 2 - Ch
PIR2:JC5205	ck: 8587	len: 160	finds: 1	! sulfur-rich protein - Chlam
PIR2:AT1529	ck: 4189	len: 303	finds: 1	! probable ribonuclease HII -
PIR2:CT1569	ck: 1406	len: 300	finds: 1	! hypothetical protein CT324
PIR2:BT3203	ck: 1180	len: 43	finds: 1	! hypothetical protein 2 - De
PIR2:SI7812	ck: 2127	len: 192	finds: 1	! hypothetical protein 192 -
PIR2:SZ2631	ck: 1875	len: 166	finds: 1	! petP protein - Rhodobacter
PIR2:BS8883	ck: 8138	len: 213	finds: 1	! transcription negative regu
PIR2:SZ6981	ck: 4069	len: 201	finds: 1	! hypothetical protein 7 (atp
PIR2:SZ7434	ck: 2111	len: 393	finds: 1	! heme binding protein precu
PIR2:SZ76728	ck: 9759	len: 187	finds: 1	! hypothetical protein - Syne
PIR2:SZ76937	ck: 9556	len: 312	finds: 1	! hypothetical protein - Syne
PIR2:SZ7451	ck: 4892	len: 333	finds: 1	! hypothetical protein sll007
PIR2:SZ7473	ck: 376	len: 156	finds: 1	! hypothetical protein sll093
PIR2:SZ7473	ck: 8425	len: 125	finds: 1	! hypothetical protein sll093
PIR2:SZ75160	ck: 4206	len: 215	finds: 1	! hypothetical protein sll163
PIR2:SZ77220	ck: 3780	len: 285	finds: 1	! hypothetical protein sll144
PIR2:SZ77428	ck: 4795	len: 79	finds: 1	! hypothetical protein srr206
PIR2:SZ41638	ck: 3600	len: 384	finds: 1	! Vans - Enterococcus faecium
PIR2:FE6937	ck: 3316	len: 224	finds: 1	! conserved hypothetical prot
PIR2:SZ39739	ck: 3644	len: 394	finds: 1	! efflux protein homolog ynfF
PIR2:SZ42365	ck: 1150	len: 208	finds: 1	! flagellar assembly protein
PIR2:CE69820	ck: 5172	len: 83	finds: 1	! hypothetical protein ynfX -
PIR2:SE6976	ck: 1724	len: 36	finds: 1	! hypothetical protein ynfK -
PIR2:SZ39656	ck: 2631	len: 252	finds: 1	! lipoteichoic acid biosynthe
PIR2:SZ72921	ck: 2352	len: 86	finds: 1	! hypothetical protein B2168
PIR2:SZ70669	ck: 7905	len: 381	finds: 1	! coenzyme F420-dependent NS,
PIR2:CT0828	ck: 2874	len: 97	finds: 1	! hypothetical protein RV0463

PIR2:D70510	ck: 2571	len: 300	finds: 1	i	hypothetical protein RV0519c	PIR2:S73488	ck: 5866	len: 385	finds: 1	i	hypothetical protein K05.or
PIR2:F70547	ck: 3126	len: 294	finds: 1	i	hypothetical protein RV0547c	PIR2:A36149	ck: 5138	len: 237	finds: 1	i	spiralin - Spiroplasma mell
PIR2:C70708	ck: 4949	len: 259	finds: 1	i	hypothetical protein RV0776c	PIR2:H69019	ck: 3145	len: 157	finds: 1	i	conserved hypothetical prot
PIR2:G70716	ck: 883	len: 282	finds: 1	i	hypothetical protein RV0953c	PIR2:D69049	ck: 1234	len: 355	finds: 1	i	conserved hypothetical prot
PIR2:D70897	ck: 9641	len: 103	finds: 1	i	hypothetical protein RV1102c	PIR2:D69065	ck: 9991	len: 249	finds: 1	i	conserved hypothetical prot
PIR2:F70958	ck: 4095	len: 212	finds: 1	i	hypothetical protein RV1377c	PIR2:G64352	ck: 6767	len: 406	finds: 1	i	2-isopropylmalate synthase
PIR2:C70763	ck: 497	len: 148	finds: 1	i	hypothetical protein RV1558	PIR2:G64456	ck: 7004	len: 167	finds: 1	i	hypothetical protein homolo
PIR2:C70541	ck: 1907	len: 170	finds: 1	i	hypothetical protein RV1577c	PIR2:D64423	ck: 9981	len: 329	finds: 1	i	hypothetical protein Mj0988
PIR2:A70639	ck: 9160	len: 181	finds: 1	i	hypothetical protein RV1957	PIR2:D64452	ck: 2920	len: 299	finds: 1	i	hypothetical protein Mj1221
PIR2:B70758	ck: 1836	len: 317	finds: 1	i	hypothetical protein RV1996	PIR2:F69386	ck: 9110	len: 437	finds: 1	i	acyl-CoA dehydrogenase, sho
PIR2:A70941	ck: 7719	len: 239	finds: 1	i	hypothetical protein RV2018	PIR2:G69440	ck: 1399	len: 136	finds: 1	i	conserved hypothetical prot
PIR2:D70943	ck: 7125	len: 324	finds: 1	i	hypothetical protein RV2037c	PIR2:A69470	ck: 6749	len: 61	finds: 1	i	hypothetical protein AF1762
PIR2:H70863	ck: 4861	len: 419	finds: 1	i	hypothetical protein RV2449c	PIR2:D69509	ck: 4730	len: 270	finds: 1	i	hypothetical protein AF2077
PIR2:E70572	ck: 8454	len: 273	finds: 1	i	hypothetical protein RV2622	PIR2:T08352	ck: 7002	len: 148	finds: 1	i	hypothetical protein H1537
PIR2:F70572	ck: 9524	len: 297	finds: 1	i	hypothetical protein RV2623	PIR2:S73242	ck: 4681	len: 291	finds: 1	i	hypothetical protein 38 - r
PIR2:G70965	ck: 4149	len: 177	finds: 1	i	hypothetical protein RV2651c	PIR2:S20471	ck: 4134	len: 86	finds: 1	i	class V zygote-specific pro
PIR2:E70885	ck: 9649	len: 308	finds: 1	i	hypothetical protein RV2859c	PIR2:S25969	ck: 7679	len: 69	finds: 1	i	hypothetical protein 69 - 1
PIR2:F70669	ck: 5693	len: 418	finds: 1	i	hypothetical protein RV2953	PIR2:T01551	ck: 1750	len: 289	finds: 1	i	hypothetical protein A_1600
PIR2:B70651	ck: 3649	len: 118	finds: 1	i	hypothetical protein RV3073c	PIR2:T05405	ck: 4069	len: 222	finds: 1	i	hypothetical protein F10M6
PIR2:D70645	ck: 2556	len: 332	finds: 1	i	hypothetical protein RV3131	PIR2:T00559	ck: 9966	len: 296	finds: 1	i	hypothetical protein F1216
PIR2:F70951	ck: 8567	len: 355	finds: 1	i	hypothetical protein RV3200c	PIR2:T01617	ck: 1414	len: 392	finds: 1	i	hypothetical protein F19F24
PIR2:C70561	ck: 48	len: 350	finds: 1	i	hypothetical protein RV3626c	PIR2:T06062	ck: 59	len: 169	finds: 1	i	hypothetical protein F19H22
PIR2:H70850	ck: 8984	len: 254	finds: 1	i	hypothetical protein RV3912	PIR2:T04272	ck: 5780	len: 349	finds: 1	i	hypothetical protein F20B18
PIR2:A70565	ck: 5027	len: 226	finds: 1	i	probable cutinase precursor	PIR2:T05639	ck: 5981	len: 445	finds: 1	i	hypothetical protein F20D10
PIR2:D70696	ck: 8388	len: 326	finds: 1	i	probable dtgp-glucose 4 - Myc	PIR2:T01144	ck: 6323	len: 272	finds: 1	i	hypothetical protein F26B6
PIR2:B70986	ck: 9024	len: 203	finds: 1	i	probable isomerase - Mycobact	PIR2:T01145	ck: 9711	len: 260	finds: 1	i	hypothetical protein F26B6
PIR2:H70760	ck: 2319	len: 202	finds: 1	i	probable lipoprotein signal pe	PIR2:T08549	ck: 4270	len: 404	finds: 1	i	hypothetical protein F27B13
PIR2:B70763	ck: 4418	len: 397	finds: 1	i	probable membrane protein - My	PIR2:T05096	ck: 9032	len: 171	finds: 1	i	hypothetical protein F28M20
PIR2:A70614	ck: 5098	len: 287	finds: 1	i	probable mmaA2 protein - Mycob	PIR2:T02589	ck: 3673	len: 381	finds: 1	i	hypothetical protein T16B24
PIR2:S13338	ck: 4079	len: 256	finds: 1	i	hypothetical protein (insertio	PIR2:T02510	ck: 8930	len: 96	finds: 1	i	hypothetical protein T19C21
PIR2:PC4110	ck: 2351	len: 137	finds: 1	i	transcription regulator homolo	PIR2:T02122	ck: 8329	len: 248	finds: 1	i	hypothetical protein T20B5
PIR2:PN0644	ck: 9954	len: 66	finds: 1	i	hypothetical protein 66 - Stre	PIR2:T02548	ck: 5743	len: 217	finds: 1	i	hypothetical protein T26B15
PIR2:S44233	ck: 6765	len: 358	finds: 1	i	strp protein - Streptomyces gl	PIR2:T05859	ck: 7427	len: 208	finds: 1	i	hypothetical protein T29A15
PIR2:S17674	ck: 4817	len: 291	finds: 1	i	serine-type D-Ala-D-Ala carbox	PIR2:T01365	ck: 1733	len: 175	finds: 1	i	hypothetical protein T29F13
PIR2:SL7717	ck: 5969	len: 280	finds: 1	i	rRNA methylase - Streptomyces	PIR2:T08993	ck: 2816	len: 364	finds: 1	i	hypothetical protein T32A16
PIR2:C64251	ck: 2836	len: 385	finds: 1	i	hypothetical protein homolog M	PIR2:T06094	ck: 7672	len: 178	finds: 1	i	hypothetical protein T5017
PIR2:I64230	ck: 1048	len: 248	finds: 1	i	sensory rhodopsin II transduce						

PIR2:T00988	ck: 916	len: 241	finds: 1	hypothetical protein T9J22.22
PIR2:T08191	ck: 2415	len: 415	finds: 1	probable translation elongation factor
PIR2:T09662	ck: 3889	len: 101	finds: 1	multicatalytic endopeptidase
PIR2:S23774	ck: 5889	len: 402	finds: 1	triose phosphate/3-phosphoglycerate kinase
PIR2:S22500	ck: 9347	len: 256	finds: 1	embryonic protein ecp31 - carn
PIR2:T16984	ck: 4925	len: 165	finds: 1	transcription factor homolog
PIR2:S47086	ck: 1305	len: 263	finds: 1	pir7a protein - rice
PIR2:S14884	ck: 7397	len: 146	finds: 1	hypothetical protein 7 - yeast
PIR2:S43448	ck: 5386	len: 215	finds: 1	FUN54 protein precursor - yeast
PIR2:S19425	ck: 1969	len: 317	finds: 1	hypothetical protein YCR015c
PIR2:S70131	ck: 1796	len: 365	finds: 1	hypothetical protein YDR273w
PIR2:S69699	ck: 3355	len: 374	finds: 1	hypothetical protein YDR415c
PIR2:S50587	ck: 5798	len: 128	finds: 1	hypothetical protein YER084w
PIR2:S64117	ck: 6897	len: 107	finds: 1	hypothetical protein YGL109w
PIR2:S46801	ck: 5569	len: 429	finds: 1	hypothetical protein YKL027w
PIR2:S64829	ck: 7470	len: 336	finds: 1	hypothetical protein YLR007w
PIR2:S68481	ck: 4630	len: 254	finds: 1	hypothetical protein YLR165c
PIR2:S69848	ck: 7663	len: 105	finds: 1	hypothetical protein YML101c
PIR2:S77699	ck: 7370	len: 238	finds: 1	inner cell wall mannoprotein
PIR2:A40550	ck: 166	len: 441	finds: 1	peroxisomal membrane protein
PIR2:S67699	ck: 4697	len: 193	finds: 1	probable membrane protein YNL1
PIR2:S60916	ck: 7226	len: 139	finds: 1	probable membrane protein YNL2
PIR2:S66918	ck: 8726	len: 157	finds: 1	probable membrane protein YOR0
PIR2:JT0589	ck: 609	len: 331	finds: 1	pectinesterase (EC 3.1.1.11)
PIR2:S03833	ck: 3269	len: 319	finds: 1	hypothetical protein 1 - chest
PIR2:A34051	ck: 7750	len: 201	finds: 1	NADH dehydrogenase (ubiquinone)
PIR2:S16556	ck: 4238	len: 289	finds: 1	hypothetical protein 1 - Podos
PIR2:S16557	ck: 4187	len: 289	finds: 1	hypothetical protein 2 - Podos
PIR2:JN0451	ck: 2342	len: 312	finds: 1	phosphoribosylanthranilate isc
PIR2:B44418	ck: 3830	len: 412	finds: 1	surface antigen - Trypanosoma
PIR2:T14622	ck: 9305	len: 425	finds: 1	hypothetical protein - Trypano
PIR2:T15458	ck: 2648	len: 331	finds: 1	hypothetical protein C08A9.7
PIR2:T15482	ck: 121	len: 217	finds: 1	hypothetical protein C10A4.3
PIR2:T15182	ck: 9118	len: 150	finds: 1	hypothetical protein C18E3.4
PIR2:T15715	ck: 4709	len: 280	finds: 1	hypothetical protein C30G12.2
PIR2:T15729	ck: 770	len: 304	finds: 1	hypothetical protein C31H1.6
PIR2:T16486	ck: 1289	len: 374	finds: 1	hypothetical protein F56D3.1
PIR2:T16528	ck: 3866	len: 419	finds: 1	hypothetical protein K02F3.
PIR2:S40941	ck: 8579	len: 219	finds: 1	hypothetical protein ZK632.
PIR2:S44857	ck: 5825	len: 170	finds: 1	PAR2.1 protein - Caenorhabd
PIR2:S44865	ck: 399	len: 336	finds: 1	R0SD3.9 protein - Caenorhab
PIR2:S06613	ck: 4664	len: 196	finds: 1	chorion protein sl9 - fruit
PIR2:S69241	ck: 7903	len: 298	finds: 1	Dreg-5 protein - fruit fly
PIR2:A24254	ck: 68	len: 255	finds: 1	ecdysone-induced protein -
PIR2:S78042	ck: 1000	len: 351	finds: 1	Ig mu chain C region, membr
PIR2:S69288	ck: 5524	len: 102	finds: 1	early lactation protein pre
PIR2:T14143	ck: 3550	len: 68	finds: 1	ATPase subunit 8 - oranguta
PIR2:JC5761	ck: 447	len: 225	finds: 1	cytokine-inducible SH2 prot
PIR2:S43928	ck: 3434	len: 310	finds: 1	estradiol 17beta-dehydrogen
PIR2:JE0175	ck: 3945	len: 317	finds: 1	friezled protein-1b - human
PIR2:T77373	ck: 9167	len: 22	finds: 1	gene N-ras protein - human
PIR2:T14788	ck: 6935	len: 112	finds: 1	hypothetical protein DKF25
PIR2:T08722	ck: 6844	len: 381	finds: 1	hypothetical protein DKF25
PIR2:T00702	ck: 4343	len: 196	finds: 1	hypothetical protein F25965
PIR2:A44478	ck: 3379	len: 373	finds: 1	probable cell growth or dif
PIR2:B44478	ck: 908	len: 337	finds: 1	probable cell growth or dif
PIR2:A61188	ck: 916	len: 358	finds: 1	probable transcription fact
PIR2:S63679	ck: 1615	len: 65	finds: 1	signal transducer and activ
PIR2:A41795	ck: 8652	len: 391	finds: 1	somatostatin receptor 1 - h
PIR2:SC5627	ck: 178	len: 225	finds: 1	STAR induced STAR inhibitor
PIR2:S41063	ck: 8093	len: 183	finds: 1	translocin-associated prote
PIR2:A36679	ck: 8392	len: 183	finds: 1	signal sequence receptor be
PIR2:A56446	ck: 3609	len: 268	finds: 1	Ig heavy chain V region (3H
PIR2:B61188	ck: 9720	len: 265	finds: 1	SC1 protein - mouse (fragme
PIR2:S63681	ck: 840	len: 64	finds: 1	signal transducer and activ
PIR2:C41795	ck: 8110	len: 391	finds: 1	somatostatin receptor 1 - m
PIR2:S66270	ck: 7215	len: 457	finds: 1	kynurenine--oxoglutarate tr
PIR2:S41286	ck: 8455	len: 223	finds: 1	lactenin - rat
PIR2:A39297	ck: 9240	len: 391	finds: 1	somatostatin receptor - rat
PIR2:SC4294	ck: 404	len: 274	finds: 1	ammonia monooxygenase (EC 1
PIR2:C75042	ck: 9304	len: 308	finds: 1	aspartate carboxyltransfer
PIR2:D71333	ck: 4876	len: 477	finds: 1	conserved hypothetical prot
PIR2:G71287	ck: 2410	len: 365	finds: 1	conserved hypothetical prot

```

PIR2:J00857 ck: 6140 len: 169 finds: 1 i hypothetical 18.1K protein - B
PIR2:T10288 ck: 562 len: 298 finds: 1 i hypothetical protein 19 - Orgy
PIR2:G72757 ck: 9964 len: 175 finds: 1 i hypothetical protein APE0052
PIR2:C72759 ck: 3453 len: 114 finds: 1 i hypothetical protein APE0064
PIR2:A72773 ck: 7998 len: 129 finds: 1 i hypothetical protein APE0172
PIR2:C72775 ck: 1302 len: 100 finds: 1 i hypothetical protein APE0189
PIR2:C72784 ck: 9325 len: 313 finds: 1 i hypothetical protein APE0258
PIR2:H72721 ck: 4962 len: 168 finds: 1 i hypothetical protein APE0313
PIR2:H72731 ck: 2984 len: 172 finds: 1 i hypothetical protein APE0393
PIR2:G72736 ck: 2336 len: 180 finds: 1 i hypothetical protein APE0429
PIR2:D72737 ck: 7680 len: 120 finds: 1 i hypothetical protein APE0434
PIR2:A72741 ck: 338 len: 160 finds: 1 i hypothetical protein APE0458
PIR2:G72640 ck: 9276 len: 147 finds: 1 i hypothetical protein APE0559
PIR2:A72651 ck: 639 len: 119 finds: 1 i hypothetical protein APE0638
PIR2:D72672 ck: 4161 len: 129 finds: 1 i hypothetical protein APE0802
PIR2:E72679 ck: 8681 len: 234 finds: 1 i hypothetical protein APE0857
PIR2:F72685 ck: 9005 len: 228 finds: 1 i hypothetical protein APE0902
PIR2:H72607 ck: 1372 len: 387 finds: 1 i hypothetical protein APE1326
PIR2:E72626 ck: 6841 len: 146 finds: 1 i hypothetical protein APE1469
PIR2:D72630 ck: 4651 len: 152 finds: 1 i hypothetical protein APE1499
PIR2:H72538 ck: 9362 len: 175 finds: 1 i hypothetical protein APE1597
PIR2:D72545 ck: 2785 len: 194 finds: 1 i hypothetical protein APE1648
PIR2:E72557 ck: 8347 len: 132 finds: 1 i hypothetical protein APE1745
PIR2:H72557 ck: 8142 len: 181 finds: 1 i hypothetical protein APE1748
PIR2:D72577 ck: 2318 len: 100 finds: 1 i hypothetical protein APE1900
PIR2:A72511 ck: 2835 len: 140 finds: 1 i hypothetical protein APE2063
PIR2:C72529 ck: 9995 len: 310 finds: 1 i hypothetical protein APE2207
PIR2:F72471 ck: 430 len: 133 finds: 1 i hypothetical protein APE2415
PIR2:C70405 ck: 5299 len: 499 finds: 1 i hypothetical protein aq_1220
PIR2:D70454 ck: 3320 len: 269 finds: 1 i hypothetical protein aq_1793
PIR2:A75196 ck: 70 len: 147 finds: 1 i hypothetical protein PAB0052
PIR2:E75204 ck: 8207 len: 179 finds: 1 i hypothetical protein PAB0107
PIR2:H75185 ck: 6767 len: 266 finds: 1 i hypothetical protein PAB2346
PIR2:D71228 ck: 1741 len: 135 finds: 1 i hypothetical protein PH0086
PIR2:D71110 ck: 6672 len: 102 finds: 1 i hypothetical protein PH0651
PIR2:E71120 ck: 8797 len: 104 finds: 1 i hypothetical protein PH0732
PIR2:H71124 ck: 5357 len: 272 finds: 1 i hypothetical protein PH0767

PIR2:F71127 ck: 8012 len: 183 finds: 1 i hypothetical protein PH0787
PIR2:F71039 ck: 7274 len: 447 finds: 1 i hypothetical protein PH1606
PIR2:F71201 ck: 6730 len: 137 finds: 1 i hypothetical protein PH1883
PIR2:D71216 ck: 2646 len: 225 finds: 1 i hypothetical protein PH1966
PIR2:G75035 ck: 3225 len: 331 finds: 1 i iron (iii) abc transporter,
PIR2:B75110 ck: 5732 len: 249 finds: 1 i lipate-protein ligase A re
PIR2:E49094 ck: 3704 len: 373 finds: 1 i methylmalonyl-CoA decarboxy
PIR2:T14222 ck: 7879 len: 344 finds: 1 i NADH dehydrogenase subunit
PIR2:S55472 ck: 8743 len: 299 finds: 1 i pectin lyase (EC 4.2.2.10)
PIR2:D72485 ck: 7054 len: 242 finds: 1 i probable high-affinity bran
PIR2:B71024 ck: 6905 len: 249 finds: 1 i probable lipase protein 11
PIR2:S72288 ck: 9584 len: 75 finds: 1 i ribosomal protein L23 - Pla
PIR2:T10349 ck: 889 len: 374 finds: 1 i very-late factor 1 protein
PIR2:S78187 ck: 3067 len: 197 finds: 1 i ymf39 protein - Reclinomona
PIR3:T11851 ck: 3584 len: 324 finds: 1 i ruvb protein - Thermus aqua
PIR4:JCS705 ck: 3534 len: 125 finds: 1 i alpha-amylase inhibitor 0.1

\\End of list
Databases searched:
NRBF, Release 62.0, Released on 30sep1999, Formatted on 26oct1999

Total finds: 561
Total lengths: 47,169,319
Total sequences: 142,080
CPU time: 05:21.97

i FINDPATTERNS on PIR: * allowing 0 mismatches
i 1 <(X) (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) (C,T,S) (R,
CCOF2T ck: 2620 len: 92 i cytochrome c2 - Rhodocyclus tenuis
1 <(X) (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W,
x(64) (P)X(S) (A)XX(V) (T) (K)XX(A)XX(V)X(12)
ADBSALAQTRGCLACHINPEKRVGPAVAGVAKKVGAGAEAKLVAKVMAGOGGWAQOLG
1 RDHUCB ck: 7216 len: 277 i carboxyl reductase (NADPH) (EC 1.1.1.184
<(X) (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W,
x(1189) (P)X(S) (A)XX(V) (T) (K)XX(V)XX(V)X(12)
MSSGTHVALVGGNGKGTALVPRICGLFSGDVLTRBDVTRGAAVQOQAGBLSRFRHOLD
1 DTKRGVHQKEGMSAAGVKKIVYVTSRTHAKRISBORRGGDKILLNACCPCGVKRI
DTKGRKATKSEEGAEIPVTLALPDAEGPFGOFVSEKRVDEM
RDALAE ck: 2829 len: 246 i acetoacetyl-CoA reductase (EC 1.1.1.36)
1 <(X) (L,I,V,M,A,P)X(P,T,S) (L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W,
x(112) (L)X(S) (L)XX(V) (T) (K)XX(I)XX(A)X(118)
MTORLAAYTGGMGGIGTAICORLAKDGRVAVAGCGPSPRRKWLFOOKALGFDFIASSEGNVA
1 VNTVSPGIATDMVKATROVDLKRIVATIPVKRLGIPETIASICAMLSSESGFST
GADSLNGGLHMG

```


1: MDDORLLSNNEQLMGRGPAFESKCSRGALYTGFSILVTLILLAGOATTAATVAFLYOQOGRUDKLL
SRHSLEOKPTDAPRPSLELEDPSGGVGTGKODLGPVPM

R3LV8 CK: 6002 Len: 132 1 ribosomal protein S8 - liverwort (Marchantia)

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{73}(L)X(S)(P)XX(L)(S)(K)XX(L)XXX(X)X{43}
x{73}(L)X(S)(P)XX(L)(S)(K)XX(L)XXX(X)X{43}

MGNDITANMTTSIRANNGIKITGYVPAINTIRNIAKILIQEGEDINDKNTDILILMLKXC

S26296 CK: 6968 Len: 151 1 ribosomal protein S13-e, cytosolic - fission

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{18}(V)X(S)(P)XX(W)(C)(K)XX(A)XXX(X)X{117}

MGRMSKRGKGIASSALPYRSPFAMCKADSDVVEQILKESKMGSPSIGVTLRDSHGIPQVRFI

A48103 CK: 718 Len: 256 1 ribosomal protein L21 precursor, chloroplast

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{39}(V)X(S)(P)XX(L)(S)(R)XX(L)XXX(A)X{201}

MASATIAFSCSLCATLKPONLNLPLPLSKPESGVYSPSLRLSLFPAARRRPFQEIPE

1: TNAAVHAAVEBOLDKDTIVFVKRRKKRRNRIGHOPITRIKITGITGYEDYPAS
TLEAVEKEEAEAEAEAVPV

JQ1298 CK: 409 Len: 340 1 annexin II type 2 - African clawed frog

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{21}(L)X(S)(P)XX(A)(T)(R)XX(L)XXX(A)X{300}

MALHEIKRKTSLFENGSSRSKSGYKAAHPDAEDAAIATKATGVDLITINILNRSN

1: GREGEDGNMNYEKIDDOARELENGYVKTDTKTLTETESLHOKYPERY
KSTSPIDIESIKKEGDELAFLNVLQCKNPLFADNLVSMGKGTDKILIRINVSCELDMAKINGEKKKYGSLAHFIC

OOFCK CK: 4358 Len: 373 1 rhodopsin - fruit fly (Drosophila melanogaster)

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{181}(L)X(S)(P)XX(W)(S)(R)XX(P)XXX(L)X{116}

MESFAVAAAOQLBPAPLNGSVVDKVPDMAHILSPYWNQFPMADMIWAKILTAIVIMIGMISWC

1: SJMCLAPAGWMSRYPEGNLTSCGIDYLERDMNRKSLIFISFYVYIPELICY
YFIIAAVSAHEKARBOAKKMNKSLSSSDAERSGKLAKVALVITILMFWMTPLYLVINGMLKFEFEGILPLNTIWGACFKSA

I55604 CK: 4750 Len: 411 1 platelet glycoprotein Ib beta chain precursor

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{94}(L)X(S)(P)XX(A)(S)(R)XX(V)XXX(A)X{301}

MPSRHTMLRPLPYVNAACSGPDRMTLVNVAAGVRLRVPLFRAGSGSLSGLEPPAIVCYLPLOR

1: GLDGAPAVVLRDVAOVADRSAYCGASLGLSLLILLAPSPRAGCPA
PSCSGTIVDCGRGTLTASLPTAPVDTTELVTGNLNLPLGLLDALPALATLALGANPWCORLVPPLRAWLAGRPERAPYRDI

QREBOB CK: 4920 Len: 306 1 oligopeptide transport system permease prot

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{185}(L)X(S)(P)XX(L)(T)(R)XX(M)XXX(L)X{105}

MNLFIRCRERLPTLITLITTSFMRMRALPSSPTGERALPEVLANIEKHYLNDPMTQFYS

1: MALSLAYASARTRGSMTEVLVLSNTRKARAGDPMRTIRFRAKPLPELV
SYMGPAFVGIITGSMVETITLPGIIGLFLNAGALNRDYSLVSLITLVALITLNAIVDLAVINDPKIRI

S67566 CK: 8415 Len: 417 1 probable membrane protein YD033c - yeast

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{196}(L)X(S)(P)XX(L)(T)(K)XX(L)XXX(A)X{205}

MLARVNLIGRRAVRYRQORLPARKADVIVAMSGVDSVAALFGEFPNTRGVYQMOMWSOS

1: PKDOSYLLSOINSTVLSLLPLIGHITKEVDRKASVGLPTAKPDSOGCFVNN
SGHGRKRNFLKHYLPSSPDITVDPOSGAKTGWHDGLSYIGQVIGISMPOADPNYQGTWSEKLRDNTNELIIVRGDNPALY

S30896 CK: 7059 Len: 297 1 virulence-associated protein mbal - Salmonella

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{23}(A)X(S)(V)XX(I)(T)(R)XX(L)XXX(I)X{256}

MDFLINKRKIKFITLMEGSSFIATSVYIRTPLSRYSIDLEBELNORQFIRKNGTILPTEF

1: GKPDINRLAGTVPVLFEGAKNNDITIVHFEORTGINTNPAPFSDNDVLSIVRL
OOGIAMLILFVRCRAGLSTGIDHALHIGVALCTSLYPTKRRTPDYRAIKILQOELKOSTF

S06670 CK: 7032 Len: 297 1 virulence-associated protein, mKac - Sal

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{23}(A)X(S)(V)XX(I)(T)(R)XX(L)XXX(I)X{258}

MDFLINKRKIKFITLMEGSSFIATSVYIRTPLSRYSIDLEBELNORQFIRKNGTILPTEF

1: GKPDINRLAGTVPVLFEGAKNNDITIVHFEORTGINTNPAPFSDNDVLSIVRL
OOGIAMLILFVRCRAGLSTGIDHALHIGVALCTSLYPTKRRTPDYRAIKILQOELKOSTF

IMECN4 CK: 6528 Len: 131 1 colicin N immunity protein - Escherichia

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{40}(A)X(S)(L)XX(E)(T)(H)XX(L)XXX(P)X{75}

MHTLLEKTIAYLSLDFGHSFNPNPLSEAFNLYHTAPLAATSLFTTHKELEIKPRSSPLRA

1: ZWECB CK: 8403 Len: 139 1 proteinase umud (Ec 3.4.21.-) - Escheric

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{73}(V)X(S)(A)XX(A)(S)(H)XX(I)XXX(A)X{50}

MLFKPADLREIVTFPLFSDLVQCGFSPADYBQRIQNLQIHPASATYFVKASGDSMID

1: H64563 CK: 9038 Len: 343 1 flagellar motor switch protein - Helicob

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{42}(L)X(S)(I)XX(I)(S)(K)XX(V)XXX(G)X{285}

MATKTRPKOAOIDEISMEKRIAILIIOGPTTCEILRHIDISISREISQIYQNGTQKQI

1: EISPOYKRVSTYKLEKSLISYKRENGEFLAVERINRGOSAKTILARISSY
DNKLGAIRKEMTFEDYIKNDIFARLEIKADKDKSLAKISTKDLTKFLNNSSRAADQFVEMQYLGAVAKIVDVAOR

BYECUI CK: 4271 Len: 130 1 unci protein - Escherichia coli

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{3}(V)X(S)(V)XX(V)(S)(R)XX(A)XXX(L)X{111}

MKNMVSIVLSRVNARKLLVOLLVAVIASGLFLSLKDPFVGSATISGGLAVFLPNVLFMIFAW

1: BBG55 CK: 173 Len: 257 1 virB8 protein precursor - Agrobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{49}(A)X(S)(A)XX(L)(S)(K)XX(R)XXX(A)X{192}

MWGDSILROIFSSAIRVDAMCEPEVAMVAESEAHEKVEAFOTARRAKSARRLKIIAAV

1: VEEHIASNDVPTSTOQIRYKRLIVVDGKMPVSTWTAIVREKVTSLPGRRLTN
PAGLVMTSTOSEDVTSNVGQAP

BBG56 CK: 2646 Len: 230 1 virB8 protein - Agrobacterium tumefaciens

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{12}(A)X(S)(A)XX(L)(S)(K)XX(R)XXX(A)X{192}

MTVARESLAHHVEVAFOTARRAKSARRLKIIAAVVAIAIIGVNAQAFATATVPSRLVAV

1: GKMFVSTWTAIVREKVTSLPGRRLTNPAGLVMTSTOSEDVTSNVGQAP

BBG58 CK: 6263 Len: 237 1 virB8 protein - Agrobacterium tumefaciens

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{129}(A)X(S)(A)XX(L)(S)(K)XX(R)XXX(A)X{192}

MKGEVSLIVARESLAHHVEVAFOTARRAKSARRLKIIAAVVAIAIIGVNAQAFATATVPSRLVAV

1: KRLLIVDGKMPMASTWTAIVREKVTSLPGRRLTNPAGLVMTSTOSEDVTSNVGQAP
HSEF

W2WJ31 ck: 9656 len: 372 1 E2 protein - human papillomavirus type 31
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(57)(V)x(P)(A)xx(V)(S)(K)xx(A)xxx(I)x(299)
 METLSORLVNCOCKLIEHENDSKRLCDHIDYWKHIRELCVMYKAREGIGISHQVVPALYSVK
 YGTGRKMEVHAGGVIFEPESVFSDEISFGIYTRKLEPTANNITTSNKTGALS
 EGVRRATISTKRPTEPERHNRHNNKLLRGDSVDSVNCVISAACACTNQRVASCPTPTPIHLKGDANILKCLRYLSKYKOLYEC

W2WJ51 ck: 1816 len: 358 1 E2 protein - human papillomavirus type 51
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(57)(V)x(P)(A)xx(V)(S)(K)xx(A)xxx(I)x(285)
 METLCHRNVCQCKLIDCELDSDKLYDQNTWTLRLYEAFAFYARENLITINHQVVPATVSK
 YGAQOEYMYGTGVTTCPEYVSTSDALSTITTEQLSNPTINPLTTCYGAKEA
 QTOQRKRORLTPEDSSITSPUSVDNTNNOIHGSGSTNGHOSATQTAFTVHLKGDINCLCKERYRTKHKGLYKNVSGTWTMTSNT

W2WJ18 ck: 5567 len: 365 1 E2 protein - human papillomavirus type 18
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(61)(V)x(P)(A)xx(I)(S)(K)xx(A)xxx(I)x(288)
 MOTPKRELSRLSCVODKLIIDHYENDSKDIDSOIQWOLIRWENAIFFAAREHGQIOTLNHOVVPAX
 SECEKXGNTGWEVHAGGVIFEPESVFSDEISFGIYTRKLEPTANNITTSNKTGALS
 VGTAKTYGOTSATRPBGHGLAEKOHCGFVNPDLGAAPTGNKRRKLSGNTPTPIHLKGDNSLKLRYLRKSHDHYRDISSTWH

S15617 ck: 734 len: 391 1 E2 protein - human papillomavirus type 2a
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(57)(A)x(P)(A)xx(V)(S)(K)xx(A)xxx(I)x(318)
 METLANLDCQCKLIDCELDSDKLYDQNTWTLRLYEAFAFYARENLITINHQVVPATVSK
 LHYGVTGWEVHAGGVIFEPESVFSDEISFGIYTRKLEPTANNITTSNKTGALS
 ESTGAGRARPTIOALCSAKAPISPPAKRORVYVGOHHRPDSITRVGSEVFCYKRSISQNSNTDPWMGDIDSVPIVHLRSDANC

QOBE5 ck: 7071 len: 366 1 integral membrane protein - salmonella hefr
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(167)(H)x(T)(L)xx(V)(S)(R)xx(P)xxx(V)x(183)
 MMRASRDETMLEWLOLLELVLELFIASVGLGSPCTFPNLDVDSLNLTLRNAKH
 SKRWVLLLEOHIPKOSLSTLIDYGRKPCVAMVSLALAEVLVSLDFPMALIGS
 FYILVSDTVLASINILFYLTTEWMTFLODYKLOFGLVFSGLLILPLVLEAYEAVVSANLHKTVAVNIMIPMCVIAM

WMN222 ck: 9828 len: 194 1 envelope-associated 22k protein - human res
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(83)(I)x(S)(I)xx(I)(T)(K)xx(A)xxx(M)x(95)
 MSRRNPCKFEIRGHCLNGKCKHSHNYFEMPPHALLVRONFMLNRLKSMDSKSIDTLSEISGAEL
 ESTVSDTNDHAKNDT

WMN2BA ck: 6337 len: 186 1 matrix glycoprotein M2 - bovine respiratory
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(83)(L)x(S)(I)xx(I)(T)(K)xx(A)xxx(M)x(97)
 MSRRNPCKFEIRGHCLNGKCKHSHNYFEMPPHALLVRONFMLNRLKSMDSKSIDTLSEISGAEL

D1NDEONE ck: 3415 len: 109 1 12k protein - potato virus M (strain Russia
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(21)(L)x(P)(P)xx(P)(T)(R)xx(L)xxx(L)x(91)
 MPTLPDPDFTRVYLSAALGVSIALVLLVLLRSTLPVVGDRDHLPHGWTYRDSGTFSVFNSPGR

VCBWVC ck: 3595 len: 237 1 coat protein - tobacco streak virus (strain
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x(74)(P)x(P)(V)xx(V)(S)(R)xx(A)xxx(L)x(147)
 MNTLIOGPDPHPSNAMSBRANNSNRCPICIDELDAMARNCBAHTVNTVSRORRNARAAYAH

RSHDFDLGKLLNLQDDRAVWCLDERREARVOLAIGWIAISKAPLMPDEFL
 VNOD

VCVERC ck: 1818 len: 351 1 coat protein - turnip crinkle virus
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(481)(A)x(P)(V)xx(V)(T)(R)xx(A)xxx(L)x(287)
 MENDPRVRKASDGAQWAKTWKQKSGSTLTSRCKOTARAAMGKILSPVQPVQKVTRLSAPVA
 SVPMGTGELTVPDSDTREVADGISDPDKYDGLKIMATYVGGADANDAGLGVRE
 YTVOLNNRNGSTSDAQIGDFAGVQGPRLVNSKTKGAGWHDCHFGTGNFSLITLYEKAPVSGLENADSPSVLGEAAAS

VHBPE6 ck: 6252 len: 149 1 major nucleocapsid protein - phage phi-6
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(118)(A)x(T)(P)xx(V)(S)(R)xx(A)xxx(S)x(115)
 MLDVVARAAVPAIESKATNPOLVSRINAAIOSKPSAIIIAVNSPVAAGLTLAIGSTG

ZGBPG4 ck: 4484 len: 177 1 gene G protein - phage G4
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(129)(V)x(S)(V)xx(V)(S)(R)xx(I)xxx(A)x(132)
 MFOKFIKRNAPINSLOLATKTPAVAAAPVLSVPLNSRSTILINATTVATTHSGLCVVRID

PDBYC ck: 6596 len: 296 1 REP2 protein - yeast (Saccharomyces cere
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(116)(A)x(S)(V)xx(V)(C)(R)xx(I)xxx(A)x(264)
 MDDEETKNTLVARARAYSWDYCRPLFIENIADVDVIDIESKRKSDLLFPGVIRPMSLST
 LVSVTQTLTVPDQQAALDGLFTFPESLKQITFNSFNLAENLOOKSLSEER
 TEVNHANTNEVPSRRTSRDNRAGAYKLONTITEGPRAVPTKRRVATVRGRKSNSTRV

B64166 ck: 1216 len: 214 1 hypothetical protein H1084 - Haemophilu
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(115)(V)x(T)(A)xx(V)(T)(R)xx(I)xxx(S)x(183)
 MNLQMKWTLITFLFAFLAVLRINAIISPLVLMQADKLFSDIOANOSKIKODPYLNT
 MLEDITKNVGIINKOGIDTLITKMOOSASOPIIFNO

E70044 ck: 4650 len: 310 1 conserved hypothetical protein yvob - Ba
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(126)(P)x(T)(M)xx(L)(S)(R)xx(I)xxx(G)x(268)
 MAKVRTDVMQENLELISGEBGINPRTWSDSRGIEIAGYFTYPPRRVQLLGKTELSEFF
 CSEIRRODDOTLVGNAPELIEHLIEHLIEIRGINVLMFGAGVNRKNKRTIYMNLE
 LMEQKQYDRDLGLEEFETKIIITETIKLIPVRPGNLAIVIEVAANNEFLRMGLNAAEOFTNKLADVIDEGEOE

669000 ck: 7972 len: 409 1 molybdenum cofactor biosynthesis protein
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(1177)(A)x(S)(A)xx(I)(T)(R)xx(V)xxx(P)x(1222)
 MGVETPFLIDLDARITGELTSELISGCTESVDLDDAGRLADDISPVDLPFPDASND
 ITRVNTISKPSVGVISTGNELEPESQGLTSPGFTFSDNSGSAKRGCSKRGK
 IYVDIGELDLALIRGLEADNVTIISGISAAGDILSEVDELGRVYIHGISTIRPGFTIVGVDDKIIVGLRFPVSAIIVM

E70191 ck: 8273 len: 337 1 conserved hypothetical protein BB0734-
 1
 1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x(1123)(A)x(S)(L)xx(L)(S)(K)xx(I)xxx(S)x(198)
 MISTEILISSQIKAAALIMGELVAFVPEFTVYVIGAGANANEDAVMIFLVYKRPINPLIVH
 IESTVVGFDLKDNYLLRGCATILKMIENELGKTYVLEKMLEKSPGMIIVH
 YKRIPIVYLFKSDNIRRYLNKDTKILITKATLKSYLEFEMWKNKNTIVFTLIEYAONLVKELVNSNNYKQILSEPLKDEELG

A71183 ck: 8100 len: 306 1 probable UDP-glucose 4-epimerase - Pyroc
 1

1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(135)(P)x(S)(P)xx(I)(T)(R)xx(A)xx(G)x(155) MKNNLIYITGAEFLSHLEALKENDVYIITDNLISGLEIIEGKVFITADYVRESIAEYISE AYAGVIFLERAIKGEPLIIEFDGKOTDFIVVDVVEANILVAKRSANGRLFN VATGKETIIELEAMKIITDMSSSSILFFPBRPGDIRSVAKIERIKLGRPRYSLEGLKETFEWTSRTO	
1	870955 len: 288 1 ctnf protein - vibrio cholerae	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(L)x(S)(P)xx(I)(C)(R)xx(I)xx(A)x(263) MCOITDLERLSPKICRSPKIKRAMSFITSRGMLTAAIISGLCYAAPLATSLEVRHANNOA LTNKSSTOVSVYGAIVGPDGVPLSGITIEGINSIMATISIVIAVMSGSL OKHAYAWNODFNEVARGALVYVPLKRAFPDPTVDLNLQVIELRLNPL	
1	A70118 len: 239 1 glycine betaine, L-proline ABC transporter	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(161)(M)x(P)(V)xx(Y)(T)(R)xx(I)xx(S)x(122) MKDPIITKIDNPFDFVDFNFSIDGFGSKITIFLYENLKNLFLVNPILFVLVCLSLFVFLK DEVTEAKSFSSNIRLIFQVQPLPSQSTIEGINSIMATISIVIAVMSGSL GATVYISIERLNFGLISGLAVYIITIDRIMOSIFIKFSTLNDHCKKKEKFRFEIYNK	
1	576375 len: 126 1 plastocyanin precursor - Synechocystis sp.	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(77)(A)x(T)(R)xx(L)(S)(H)xx(L)xx(A)x(33) MSKRFITLGLLVSSFTLSVSRMAANATVKNSSGSLVPEPSTVIRKGEVKNWNKI	
1	F70646 len: 323 1 probable quinone oxidoreductase showing sti	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(156)(L)x(P)(A)xx(A)(S)(R)xx(A)xx(I)x(151) MRAVRYTRLEGDAVEAVEEPTSAQVIEHAGVAFPDALITGRQYRPPPEPLVGAELIAGV LATVAGATDVLAEGFKDAVOELTNGRCVDTIVDPVGGDFITSLRSLLAAGRLV IGFTGEIPTVYNRLNINIDVGVGWSLTHPDALAOOWSOLERLLRSGKLPPEPVVYPLDQAAAIAIENRTAKKAVLVN	
1	S01901 len: 254 1 alcohol dehydrogenase (EC 1.1.1.1) 2 - frul	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(146)(P)x(P)(V)xx(A)(S)(K)xx(A)xx(T)x(92) MAIANKNIEFVAGLGIGFIDISREIVKSGPRNLVILDIRIENPAIAELKALNKPVTVFYPPDYVY SINPGITKTIVHKNSMIDVEPRVAELLLHPOTTLQCONFVKAIQANONGAI WRKLDGLTLEAIEMTKHDSHI	
1	S65732 len: 132 1 carbonyl reductase (NADPH) (EC 1.1.1.184)	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(66)(P)x(S)(A)xx(V)(S)(K)xx(V)xx(S)x(50) SNVFAVAVITGNKPAAGCEAAVAKLOEBGLHPLFHLDDIDQVSDRPFVAVQAEVTLKGVHEEGK	
1	JC5284 len: 277 1 carbonyl reductase (NADPH) (EC 1.1.1.184)	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(189)(P)x(S)(A)xx(V)(T)(K)xx(V)xx(S)x(177) MSDRPALVATGANKKIGFIRLVDCRKLFLVLANDESREGAANKOLOTESLSPRHQDIDK DARKGHAKEGMENSAYGVTKIGVTVLSRIVARKLNEBRREDKILNACCGWVRT DMAQPAATKSPBEGATPYLALLPFGAGPHGQFVODKVEPW	
1	S54815 len: 261 1 urdamycin polyketide ketoreductase (EC 1.1.	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(114)(L)x(S)(V)xx(M)(T)(R)xx(L)xx(G)x(111) MTOONPVAIVTGAATSGIAGVATRLGKHGVFICARTEDSVSTVAKQLIDEGLEVADGACDVR5 TGITVNAVCGYVEITPMAORVARGVAAAYTSEDALIEKFOAKIPLGRYSTPEEVA	
1	GLVGYLASDPTAASITSOALNVCGLGNF	
1	D55587 len: 261 1 griseusin polyketide ketoreductase (EC 1	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(114)(L)x(S)(V)xx(V)(T)(R)xx(L)xx(G)x(131) MSDDKVAIVATGATSGIGISVARIASGHRFVGARSERNAATVKELVAGLDVGDVAVND TGITVNAVCGYVEITPMAORVARGVAAAYTSEDALIEKFOAKIPLGRYSTPEEVA GLVGYLASDPTAASITSOALNVCGLGNF	
1	S69225 len: 262 1 nogalamycin polyketide ketoreductase (EC	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(115)(L)x(S)(V)xx(V)(T)(R)xx(L)xx(G)x(131) MPDIDKVAIVATGATSGIGISVARIASGHRFVGARSERNAATVKELVAGLDVGDVAVND KSGITVNAVCGYVEITPMAORVARGVAAAYTSEDALIEKFOAKIPLGRYSTPEEVA AGLVYITPTPADSITAOALNVCGLGNF	
1	S35196 len: 261 1 hypothetical protein 3 - Saccharopolyspo	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(114)(L)x(S)(V)xx(M)(T)(R)xx(L)xx(G)x(131) MAANTKPYALVATGATSGIGISVARIASGHRFVGARSERNAATVKELVAGLDVGDVAVND TGITVNAVCGYVEITPMAORVARGVAAAYTSEDALIEKFOAKIPLGRYSTPEEVA ANVEYLVSDDGAAVTAQALNVCGLGNF	
1	D64051 len: 242 1 3-oxoacyl-[acyl-carrier-protein] reducta	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(108)(L)x(S)(V)xx(L)(S)(K)xx(L)xx(M)x(118) MOGRKALVGTSGRIGGAIIEELSSKGAFTGATEKGAEMIASVADDKGGLVANTDKES HNQGLIYS	
1	D69648 len: 254 1 2-keto-3-deoxygluconate oxidoreductase x	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(154)(L)x(P)(A)xx(A)(S)(K)xx(V)xx(T)x(184) MGYIDPASTGKTAIVTGPDTGIGGIAKALAGACADITGTSHTSSLETQOLVEDGRIFT MASGICVNAIAPGISTANTPTPINDDEKRNEDILKRIAPACRMOANDIGTAVFLA SRASDIYVNGHILAVDGMWLR	
1	C48674 len: 268 1 tropinone reductase homolog - jimsonweed	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(162)(L)x(S)(I)xx(A)(S)(K)xx(I)xx(T)x(190) MAGREIGGDRWSURGNTALVTGKIGIATVEELANGALVITCSRSQNDDBCELEKRR TSGIACEMAKDSIRVNAVAPWITINPTIEAACQVPSQKNIESLIGRAPKRKRGEP SEVSSIVTLCPLFTASITGOTICVDDGYTVNGFI	
1	T12020 len: 247 1 3-oxoacyl-acyl carrier protein reductase	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(113)(L)x(S)(L)xx(L)(S)(K)xx(L)xx(T)x(118) MSIOGVAIVATGATSGIGISVARIASGHRFVGARSERNAATVKELVAGLDVGDVAVND TGATVNAVCGYVEITPMAORVARGVAAAYTSEDALIEKFOAKIPLGRYSTPEEVA TLHVGSGMXY	
1	T12051 len: 244 1 3-oxoacyl-acyl carrier protein reducta	
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x(110)(L)x(P)(I)xx(M)(S)(K)xx(L)xx(M)x(118) MNEEGKALVATGATSGIGISVARIASGHRFVGARSERNAATVKELVAGLDVGDVAVND TVAPGIEITDTRALNDQRAATLSVNRAGRIGDRRELASVAVFLASPEAAVITGE	

1	507557	ck: 7659	len: 268	1	cytochrome-c oxidase (EC 1.9.3.1) chain III
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{56}(V)x(S)(L)xx(W)(T)(R)xx(V)xxx(S)x{126} MSTKIQSSASHLOVSHSWPILVFSFNIAITVLMGHISSTTDTDLGLAVTGSILMTRE 1: MSTKINIGLETSLYMHFVDLWMLFTFLVYEWGA
1					ISYFOAFEXIHAGFMSDSVGTVEFASSTGLGHJHWMGLTFLFVSFLROYQT
1					TKHNIGLETSLYMHFVDLWMLFTFLVYEWGA
1	709997	ck: 1856	len: 385	1	lipoxigenase (EC 1.13.11.12) - southern As
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{124}(R)x(P)(A)xx(I)(S)(R)xx(A)xxx(P)x{245} 1: WMHSSSCFCFALPGMPLAMKLAHVALHSGYQALSHSLRHCALIEFYIILASNRQLSAH PRESSPSESTELQEMKRETRRYGADKRDAGMDLTPDLIGILTITWOSG
1					HHANVNGQYVDAGYFPPNRTTARTCMPHDSEEMWHPKKREDTLKCFSQLQATRYKAVLDVLSNHSDEEYLGASPEYWI
1	543771	ck: 2848	len: 347	1	phosphatidylcholine desaturase (EC 1.3.1.35
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{35}(I)x(P)(L)xx(W)(S)(R)x(V)xxx(L)x{126} 1: MSTVSPSTATLEKPRNRLDLDTIPRYEITNLRKMSVYLSVAAYGCAALLAIPWIL FNMEAFGKQREOVRESALFVILLAGIAPVMEYTLGWGVYKFMLEPMLEGRM STFLVHHTVPELFPFSYRDMKDAIQLSGTVHCDPKRMEVYLDHIVHFMHLSGTIPSYLKRKAYASIKQNGELLYETKSWEL
1	574692	ck: 6210	len: 231	1	crRNA(m13g7)methyltransferase - Synecocyst
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{70}(V)x(S)(L)xx(L)(S)(R)xx(V)xxx(T)x{145} 1: MOPDLTLFPDEFPSPLQSLQILKALKALKAIAVSNINPRDEITDHRVNDDEPYGGGVANIKPEH YTRPVPVGLAVPVLGSHNOAIAOMRLEOQEEERTQRRPDLMOQMORQSP
1	A71119	ck: 9980	len: 308	1	probable aspartate carbamoyltransferase cat
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{17}(V)x(S)(I)xx(P)(S)(R)xx(I)xxx(L)x{285} 1: MEMGRDVISIIRDSKEDIYVLSIAERLEKEREKGOLEYKGLIATLTFEFPSTRTLRSPESAM AYVDLILISPELLIPMKRYHVEITLPRGMKYEITLKEVIGELDYLVYRIORE RPEDDETLKRGKSGVKNKLEWVKSLSIMHPRLPRVDEIHPEVDKTHAIFYQVNGVPMALLAVLGI
1	A70959	ck: 1547	len: 319	1	probable aspartate carbamoyltransferase - M
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{3}(L)x(T)(A)xx(L)(S)(R)xx(A)xxx(L)x{298} 1: MPRRLTLADLSKRDARLIDDDADRQALDYGRDKIKPTLRGTYVTMYENSTIRTSFVAG NMMLDLTGAEVAVLPETLLPVGVTWPTVSHDDAPADAYDLKRYDAEKS NGFFPVSREYSRYGLTEROAMLPGHAVLHPRGVPRKMETTSVADSOSAVLQOVSNQVYRAVLYPHVLVAGQADAGEGAA
1	C70815	ck: 556	len: 403	1	probable beta-ketosidyl CoA thiolase - Myc
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{155}(I)x(T)(I)xx(F)(S)(R)xx(V)xxx(A)x{232} 1: MSEBIEFYALITPRGKQKNGSLHEVPLSLVGLIDELRKRPDLIDENLISDVILGCVSPVDOG AAEAGSGVFAFVAVPDONGLLIDDEHMRDITREKLAKPAEFGALAG FDVAVLQKTHWEKINHVTGNSGIVDCAALYIGSAAGKLOGLTRARIVATASGADPVIMLTGPTPATRYLDARGLTVDI
1	T12981	ck: 8558	len: 452	1	glucuronosyl transferase-like protein - Ara
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{187}(L)x(P)(L)xx(W)(C)(R)xx(V)xxx(T)x{249} 1: MERVRKRITVYPAVAGOHVTPMMLQKALOSGFLITVAGROFMOISSLOHPGDFVTIYES KDLPTGSGPDLLEIMPRVYNAKRTASAVININAGLESLSLWLOELGIVYV LGLHLITASSFGSLIOEMSGITELNKRKRSVYISLTQKAMHETKEMLEXAMGLINSQPFVWVIRPGSVAGFEMIELDEEVIR
1	140966	ck: 7621	len: 323	1	streptomycin/spectinomycin adenyllyltransfer
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{187}(L)x(P)(L)xx(W)(C)(R)xx(V)xxx(T)x{249} 1: MERVRKRITVYPAVAGOHVTPMMLQKALOSGFLITVAGROFMOISSLOHPGDFVTIYES KDLPTGSGPDLLEIMPRVYNAKRTASAVININAGLESLSLWLOELGIVYV LGLHLITASSFGSLIOEMSGITELNKRKRSVYISLTQKAMHETKEMLEXAMGLINSQPFVWVIRPGSVAGFEMIELDEEVIR
1	507557	ck: 7659	len: 268	1	cytochrome-c oxidase (EC 1.9.3.1) chain III
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{56}(V)x(S)(L)xx(W)(T)(R)xx(V)xxx(S)x{126} 1: MSTKIQSSASHLOVSHSWPILVFSFNIAITVLMGHISSTTDTDLGLAVTGSILMTRE 1: MSTKINIGLETSLYMHFVDLWMLFTFLVYEWGA
1					ISYFOAFEXIHAGFMSDSVGTVEFASSTGLGHJHWMGLTFLFVSFLROYQT
1					TKHNIGLETSLYMHFVDLWMLFTFLVYEWGA
1	709997	ck: 1856	len: 385	1	lipoxigenase (EC 1.13.11.12) - southern As
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{124}(R)x(P)(A)xx(I)(S)(R)xx(A)xxx(P)x{245} 1: WMHSSSCFCFALPGMPLAMKLAHVALHSGYQALSHSLRHCALIEFYIILASNRQLSAH PRESSPSESTELQEMKRETRRYGADKRDAGMDLTPDLIGILTITWOSG
1					HHANVNGQYVDAGYFPPNRTTARTCMPHDSEEMWHPKKREDTLKCFSQLQATRYKAVLDVLSNHSDEEYLGASPEYWI
1	543771	ck: 2848	len: 347	1	phosphatidylcholine desaturase (EC 1.3.1.35
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{35}(I)x(P)(L)xx(W)(S)(R)x(V)xxx(L)x{126} 1: MSTVSPSTATLEKPRNRLDLDTIPRYEITNLRKMSVYLSVAAYGCAALLAIPWIL FNMEAFGKQREOVRESALFVILLAGIAPVMEYTLGWGVYKFMLEPMLEGRM STFLVHHTVPELFPFSYRDMKDAIQLSGTVHCDPKRMEVYLDHIVHFMHLSGTIPSYLKRKAYASIKQNGELLYETKSWEL
1	574692	ck: 6210	len: 231	1	crRNA(m13g7)methyltransferase - Synecocyst
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{70}(V)x(S)(L)xx(L)(S)(R)xx(V)xxx(T)x{145} 1: MOPDLTLFPDEFPSPLQSLQILKALKALKAIAVSNINPRDEITDHRVNDDEPYGGGVANIKPEH YTRPVPVGLAVPVLGSHNOAIAOMRLEOQEEERTQRRPDLMOQMORQSP
1	A71119	ck: 9980	len: 308	1	probable aspartate carbamoyltransferase cat
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{17}(V)x(S)(I)xx(P)(S)(R)xx(I)xxx(L)x{285} 1: MEMGRDVISIIRDSKEDIYVLSIAERLEKEREKGOLEYKGLIATLTFEFPSTRTLRSPESAM AYVDLILISPELLIPMKRYHVEITLPRGMKYEITLKEVIGELDYLVYRIORE RPEDDETLKRGKSGVKNKLEWVKSLSIMHPRLPRVDEIHPEVDKTH

1	IS1665	ck: 4724	len: 360	1	arginase 3 - African clawed frog
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{90}(P)x(T)(V)xx(A)(C)(K)xx(A)xxx(G)x{254} MSIRSFNFKLLKKQVNIIRKQKCSHSAVIGAPSKQKRRGVEHGPALRSAGLIDRLSNLGN PPIPGFSWAKPCLSKSDIVYIGLRDLDPAEQFIILKNYINISYSMRHADCMGIRKVM EXTFDOLLGRDRPIHLSFDIADDPALAPATGPVIGGLYRGGVYITIEIHNTGMLSDALVEVNPVLATYSEEVKATANLAVDVI			
1	JS0609	ck: 8425	len: 140	1	biastictidin-S deaminase (EC 3.5.4.23) - Bac
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{95}(V)x(S)(P)xx(M)(C)(R)xx(I)xxx(A)x{129} MKTFNISQODDELVEVATEKITMLVEDKHHVGAIRKTEIISAVHIEAYIGRVVCAEALAIIG			
1	S37304	ck: 2135	len: 335	1	spal protein - Salmoneilla typhimurium (frag
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24}(V)x(P)(P)xx(A)(S)(R)xx(V)xxx(L)x{295} DPLGKIVERTFEVAPISSEERIVDAPSVASRGVREPLITGVRAIDGLITCGVGQRMGIFASAG YPAVSENLPLLEKPGATSEGSTIAFTVVLSEEDADPADIRSIIDGHLYS RKLAGQGHYPAIDVLSKYSRVFGQVTPTHAEQASAVRKLMTLELQFLDIEGKRGGENINDRAMQWRDLSLKAMLCQPVAYQYSSE			
1	S01397	ck: 4392	len: 127	1	H+-transporting ATP synthase (EC 3.6.1.34)
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{68}(V)x(T)(L)xx(L)(S)(R)xx(L)xxx(A)x{43} MGNLQAMFMRQVRIILYLAITYLIGFGFTPYKTVFSLITGTSISLWMYNTWKIEFGQAVV			
1	S22348	ck: 1686	len: 168	1	H+-transporting ATP synthase (EC 3.6.1.34)
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{141}(V)x(T)(A)xx(A)(T)(R)xx(I)xxx(I)x{11} MLPALLRNPGLRGLRVHRAVAAENAAAPAAASGPNOMSFIFASTOVFFGANGVAVROYDPTLTGA			
1	D72121	ck: 2102	len: 209	1	ATP synthase chain D - Chlamydia pneumoniae
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{119}(L)x(T)(M)xx(A)(S)(K)xx(V)xxx(V)x{74} MSVOYKLTENSRFLERKOKLARLQIYLPILKLLKALQAEVONAVKDAECCDKDYQAYERITVAFAR			
1	VFLSDRSITDVGVKMAKKKIELKARDECV	1:			
1	S31498	ck: 4433	len: 181	1	ribulose-bisphosphate carboxylase (EC 4.1.1
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{5}(V)xx(S)(A)xx(A)(T)(R)xx(V)xxx(S)x{160} MASSIVSSAAVATRSNVQAQSMVAPFTLKSASAPVTKKNNVNDIISLANSNGSRVACMQVMPPIN			
1	PGCY	1:			
1	A71295	ck: 255	len: 357	1	probable alanine racemase (alr) - syphilis
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{184}(A)x(S)(A)xx(L)(C)(H)xx(A)xxx(M)x{157} MSRTRARVCLPVKAAAYAGHAGCDVQAOLSCGVSHFVAVACQEAQASRAAGVAPILCLSTPRAEF			
1	SIPLVHAANSALLCHPRAHFDVMPGGLAAGVAPESVHPAARSLVPELVETVOV	1:			
1	RAIKITIPAGAYSVORLWRAHTEHVGILPISGYADGVKALSPGLQYIGKMYPPVYGAALCMQCQVVDLGTPLRYTVYGRVTLFPGPOC	1:			
1	G70553	ck: 8877	len: 268	1	probable echA10 protein - Mycobacterium tub
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{30}(P)x(L)(L)xx(L)(T)(K)xx(L)xxx(A)x{122} MSNYRDEFTIYVPGIAVTLAAGSVLTITDREPSISLTKPVLAAQMAALISGAADPRVXKVVRLGGA			
1	AAEALSMGIVASVAPADDEFDQVNDKLSRLSLAGAAALAIATKNAINATITELAPST	1:			
1	LTRELDGALLLRKTDDEAFEGTAFQGRRTPEFTGR	1:			
1	H69457	ck: 4102	len: 322	1	ornithine cyclodeaminase (arcB) homo1og
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{178}(V)x(P)(A)xx(A)(S)(R)xx(V)xxx(T)x{128} MNTLLTQEEVSESLISMDANNAVEAEAFRLVALCKAQMPPRYVLEEFKGDLRAMPALHMGYAG			
1	SVQPAEASRCQDVLTPTPSKRPVAKAEWEEGTTHVAGLADGPKGEDELILKK	1:			
1	AKIVVDLEQAKHGGEINVAVSKGVIEVDVHATIGEIVAGLKGRESDSEITIFDSTGLAIDQVAAVAKVYENALSKRVGSKIK	1:			
1	S75089	ck: 8968	len: 338	1	UDP-glucose 4-epimerase - Synechocystis
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{141}(P)x(S)(P)xx(A)(S)(K)xx(V)xxx(L)x{181} MATQOTIIVTGGAGYIGSHGVLAIOQAQFDVLTIDNLSYGHKELVQPLGVELVYVGHGDRQKL			
1	GADPGRLGEDNHPETHILIPALITLAKQSRPOLSVEGTQVDPLDGTALRDYTHVCD	1:			
1	LAIAHVGLQYLLSGESNIFNLGNGSFVRQYIEFAKVITGLDIPYQLCPRRPDARILVGSSAKAREILGNMPOYPDLTHTII	1:			
1	H69105	ck: 3451	len: 336	1	dTDP-glucose 4,6-dehydratase - Methanoba
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{142}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{178} MEKIIIVTGGAGYIGSNFIRYMLQHPYHIIIMDLALTYCGHLENLRGEDEPRYTFVRSITDR			
1	YGPYOFPEKLIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	EYVNIQNSERSEENIEVELIVRELKDESDILREVEDRPHDRRYADASKIRNELGMPLYSFEEGIRETIRMYIDNDRMWNENIK	1:			
1	D69290	ck: 128	len: 332	1	dTDP-glucose 4,6-dehydratase (rfdB) homo
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{142}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{174} MRLIVTGGAGYIGSNFVYALAHQVLEKLNIDALCKYNSNENLKDLAEDERSYFTIGDINDYE			
1	YGPYOFPEKLIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	EYVNIQNSERSEENIEVELIVRELKDESDILREVEDRPHDRRYADASKIRNELGMPLYSFEEGIRETIRMYIDNDRMWNENIK	1:			
1	S47045	ck: 5086	len: 346	1	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{160}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{170} MTEGKKRILVTGGAGYIGSNFVYALAHQVLEKLNIDALCKYNSNENLKDLAEDERSYFTIGDINDYE			
1	YGPYOFPEKLIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	EYVNIQNSERSEENIEVELIVRELKDESDILREVEDRPHDRRYADASKIRNELGMPLYSFEEGIRETIRMYIDNDRMWNENIK	1:			
1	VEDHARALYQVVTGCVGEYINISGHNERTINLEVVATICALLEELAPENRPAQVAREDLITTVQDRPGHDARYAVDAKAKIRDLGAG	1:			
1	AMORTYRLPSTVSCSNNNGPBPPEKTIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	YGPYOFPEKLIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	E70566	ck: 5640	len: 331	1	probable dTDP-GLUCOSE 4 - Mycobacterium
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{142}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{173} MRLIVTGGAGYIGSNFVYALAHQVLEKLNIDALCKYNSNENLKDLAEDERSYFTIGDINDYE			
1	YGPYOFPEKLIPLMTALNENKPLPVAGDQMNVRDWHVLDHCRAVDLVLHRCRGV	1:			
1	RTYLISSEGERDNLTVLRTILRLMDRDPDFDHVTDVROVGHDLRYAIDPSTLDELQWAFKHATDFEEGLRTTIDWYRDNESMMWRPLD	1:			
1	G70415	ck: 2675	len: 321	1	nucleotide sugar epimerase - Aquifex aeo
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{146}(P)x(S)(P)xx(A)(S)(K)xx(A)xxx(V)x{159} MMLIVTGGAGYIGSNFVYALAHQVLEKLNIDALCKYNSNENLKDLAEDERSYFTIGDINDYE			
1	YFTVYVGLAGRPDMVNFVFIYKTKLKGPIKRYKQSGKQRPDTYVDAEVTALVNLN	1:			
1	KGYELINVGNKPRALKELELIEIKTKYGRKXYEYDFHKAQDRDQWADITAKKLLGMGPRTSLDEBGVKTVEWFEENMWMYVMD	1:			
1	E69750	ck: 3388	len: 249	1	glucosamine-6-phosphate isomerase homo1o
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{184}(I)x(T)(I)xx(E)(S)(K)xx(V)xxx(S)x{149} MKILIAHEYDEKLSAIIIRKQDQAKDRDVLGATGSTPVGLYKOLISDYOAGEIDFSKVTYTT			
1	FLAISMGIKTITMESKHIIVLLASGEERKADAIQMAEGPVTYDPAISILQKHNHVTV	1:			
1	IADYKAAOKKKSASF	1:			

1	J00472	ck: 2831	len: 135	1	T-cell receptor beta chain (BTB4) - bovine
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(124)(L)x(S)(P)xx(V)(S)(K)xx(A)xxx(I)x(97)			
1		MISQHRQDSLGAGRKQALGLALSSLSPPAAVSKGASVTEICRALDFQATVFWYRQEPKGLMLM			
1	J00473	ck: 3911	len: 136	1	T-cell receptor beta chain (BTB1) - bovine
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(123)(L)x(S)(P)xx(V)(S)(K)xx(A)xxx(I)x(97)			
1		MISNRQDSLGATRRKALGLALSSLSPPAAVSKGASVTEICRALDFQASMEWYRQEPKGLMLM			
1	B53250	ck: 6952	len: 353	1	class I histocompatibility antigen BOLA-AM1
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(143)(V)x(T)(H)xx(I)(S)(K)xx(H)xxx(G)x(188)			
1		LLSGVAVLETRAGSMTSTAVSRDELGPRLLEVYDDTQFVPSDADPMPREPARMY			
1		CYEMLRRLYENGKDTLLRADPPRAHVTTRHPISGREVTLKCMALGFYPERISLTWQR			
1		NGEDQTDMEVETRRPSGDNFQKMAALLVPSGEQKRYCOVHEGLOEPLTKWEPPOSFLTLGIIIVGLVLLVTGAVAGVVIC			
1	I50609	ck: 8972	len: 207	1	T-cell surface glycoprotein CD8 beta chain
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(140)(M)x(T)(L)xx(L)(T)(K)xx(M)xxx(P)x(51)			
1		MARPMLTMCLOLPGEFTNLSSQFPGYILTKTNSTTEIVCPMKGEHTGVYWMNOGRHFEPL			
1		GALLLLSLPTIRREYRLRLRMVRAHR			
1	I39516	ck: 7310	len: 257	1	ribosomal protein S3 - Achaetoplasma axanthu
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(92)(I)x(S)(L)xx(L)(T)(K)xx(I)xxx(P)x(149)			
1		MGQKVNPIGFRVGIIRMDSKWYADKTIYPALIKEDAVIRKFLNKYNNAAVSHYIERIKELAKY			
1		PLHTLRADVEYATAEQTYGILGIVMITHGSLPGSHHEELKROSSASSNNG			
1		GGRRPSKGRPSRSDAATBEGN			
1	S77489	ck: 8926	len: 133	1	ribosomal protein S8 - Synechocystis sp. (S
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(74)(I)x(T)(L)xx(V)(S)(K)xx(L)xxx(S)x(43)			
1		MASTDITSDMLTRIRNACAVRHSTTQVPTTKMLTIKALYKSEGFIEDYSETEGINKMLVLTLYK			
1	T01947	ck: 4063	len: 252	1	hypothetical protein Fllo4.1 - Arabidopsis
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(77)(I)x(T)(P)xx(V)(S)(R)xx(A)xxx(A)x(229)			
1		MPTRIRIAGTGPESGRDPAIRAPAFEFSEWITVIFPGQSGNAYGKLIGDGPATPAGLVAASLSHA			
1		YDTIFIGSNGHEPLPSNDF			
1	A39234	ck: 9791	len: 371	1	opsin - bluebottle fly (Calliphora vicina)
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(179)(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x(176)			
1		MERYSTPLIGSFALNNGSVTKVTPDMAHLVHPYWNQFPAPAEFWAKFLAYVWLATATISWGN			
1		MTLAPVFGMSRYVPEGNLTSCGIDYLERQWNPSTYIEFISIVYVLPFLICYSTW			
1		FIILAAVSAHEKARBOAKKMYKSLRSSBDADKSAEGKLAKALVTISLMEWAMPYTIITLIGLFEKYEGLTPIINTIGACFAKSAAC			
1	S40691	ck: 7181	len: 374	1	opsin rh1 - fruit fly (Drosophila pseudoobs
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(1182)(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x(176)			
1		MTSFAVATOLGQROFAAPSNGSVYDKVTPDMAHLISPYNDQFAPAMPYAKILITAYMITIGMSIC			
1		TIWCCLAPVFGMSRYVPEGNLTSCGIDYLERQWNPSTYIEFISIVYVLPFLICY			
1	JC4304	ck: 2863	len: 355	1	orphan G-protein-coupled receptor - huma
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(80)(V)x(T)(L)xx(W)(T)(H)xx(I)xxx(G)x(259)			
1		MDQFESVLENFEDYDLAECYIGDIVGTFVGLFISFVPAIDLVGNLVLVALNNSKRPK			
1		DYDEVLOEIMPVLRNENFELGFLPLILMSYCYFRIIOTLFQCKNHRKAKAIKI			
1		LLVYVIFEFLEWTPYNNMIFLETKLYDEFPSCDKMRDLALSVETVAFSHCCLNLIYAFAGEKFRRLYLHLGKCLAVLCGR			
1	I58186	ck: 4478	len: 354	1	hypothetical G-protein coupled receptor
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(81)(V)x(T)(L)xx(W)(T)(H)xx(I)xxx(G)x(257)			
1		MPTSPFLDENFEDYDSDAECYIGDIVGTFVGLFISFVPAIDLVGNLVLVALNNSKRS			
1		GDPYEVLOEIMPVLRNENFELGFLPLILMSYCYFRIIOTLFQCKNHRKAKAIKI			
1		ILLVYVIFLEWTPYNNMIFLETKLYDEFPSCDKMRDLALSVETVAFSHCCLNLIYAFAGEKFRRLYLHLGKCLAVLCG			
1	A29667	ck: 8980	len: 60	1	pulmonary surfactant protein B - bovine
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(27)(A)x(T)(V)xx(V)(C)(H)xx(P)xxx(G)x(17)			
1		FPPIPTCMLTRILIRIKRIQAVIPRGVLAIVAOVCHVPLVGLIIQOLVIEYSVILXTD			
1	S22494	ck: 429	len: 277	1	rRNA N-glycosidase (EC 3.2.2.22) alpha-1
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(57)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(204)			
1		MKRFTVILIALIFAASTVADADVAFSSGSSSYSGKFIQDLKRALPSNGTYNNITLLSSASG			
1		AAASRRYIEGQIIEISKNQVPSLATISLEMSALSQIOLOQNTNGTKTPTV			
1		ITDDKQORVEITNVTSKVYTKNIOLLNKKOVNAFDDVDSAKH			
1	S23519	ck: 9157	len: 278	1	beta-luifin - smooth loofah
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(59)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(203)			
1		MNRFTSLILILIAFTVAGVADVAFSSGSSSYSGKFIQDLKRALPSNGTYNNITLLSSASG			
1		TTAASRKYIEGQIIEISKNQVPSLATISLEMSALSQIOLOQNTNGTKTPTV			
1		VVIIDNKGORVEITNVTSKVYTKNIOLLNKKOVNAFDDVDSAKH			
1	JN0108	ck: 7681	len: 250	1	luifin-b - smooth loofah
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(38)(L)x(S)(A)xx(A)(S)(R)xx(L)xxx(S)x(196)			
1		ANYFSISGDSKSYKFTTALRKALPSEKSNIPDLLPSASGASRYITLMQLSNYDAKATIM			
1		NEVSPPAALSLEREASLSLQIOLOQNTNGAFRTPVITIDNKGORVEITNLTASKV			
1		QINDVNSKILLNKKONIA			
1	S65052	ck: 5285	len: 214	1	pistil-specific protein stsl4 precursor
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(43)(L)x(T)(A)xx(L)(T)(H)xx(A)xxx(P)x(155)			
1		MEVLTSMACLVITVITVITDEKREKELAVRNKNTLDFQFLTLTASSLHISAGTYPPPPP			
1		KKSIELGCAORCYESPATLVCFNPNENGTGERPY			
1	JC4822	ck: 8148	len: 77	1	acyl carrier protein - Bacillus subtilis
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
1		x(A)x(T)(L)xx(V)(T)(K)xx(V)xxx(G)x(60)			
1		MAQTLERVTKIIVDRGLGVDEADVLTAEAFREDICADSLDVLELMELEDFDEWISDEDAE			
1	B65003	ck: 7651	len: 228	1	histidine transport system permease prot
1		<(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			

1:	KL251(A)x(T)(A)xx(F)(T)(R)xx(V)xxx(1)x(87) MLYGGVILGALVATLELAISVLAIVLIGLIGAGGKLSQNLGSLIEGVTILIRGVPDLVMI KATOLAGKSTWEPFAIVCGVILVFTIVNSGLVLEFERRYSVGVKRAADL				
B36263	ck: 5147 len: 306	1	oligopeptide transport system permease prot		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{185}(A)x(S)(I)xx(I)(T)(R)xx(M)xxx(L)x{105} MLKFTLRCLLEAPLITLITISFEMRLAAGSEPTGERLPEPVANTLAKYHLNDPIMTQYESY MYALASIVASIRARITRSGMIEVLHNSFTIRARAKGLPMRIILRHAKPALLPV SYMGPRFVGIITGSVIEITIGLPGIGLFGVNLNDYSLVLTILVGLTALIDNALVDAVIDPKIRY				
C71163	ck: 4618 len: 304	1	probable oligopeptide transport system perm		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{55}(A)x(T)(L)xx(M)(T)(R)xx(I)xxx(T)x{233} MEYIKLAFNNKFKGFGGLIAFIIFGLIPFPASDGLYEYGVKIRIASYSAKTLPPMTREX ARCVRAQTLSTLNREYVHNSKIMGVGDIPIYEDILNPMISYIFMGILGVSGAIL ASATLDFIGLGFITMVSGLVILQKALIMNALOFGMMWFIFPGILITLITSLPFINLGIEVFNRRLGV				
S64728	ck: 6546 len: 482	1	protein secretion protein xcpR - Pseudomon		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{151}(A)x(S)(A)xx(V)(S)(R)xx(V)xxx(L)x{315} MLPYRLARAGLAMPAGOGQWOLMRDSEVLOELRHAGPSALATYEPAPFEDQLAQYOGG GRITLRAAGREVDVRSVTLPGIIGERVYRVLDKQALDLNIGMPAAVHLGLRS CLARPGVILSTGPGTSGKTLTILVSLNSLNDGSKNLTVEDPVEVALIGOTAINPRAGLTFASGLRAITRODPVIMGEIRDOH				
S52163	ck: 8969 len: 331	1	sucrose specific repressor - Escherichia co		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24}(A)x(S)(V)xx(A)(T)(R)xx(V)xxx(I)x{291} MASLKDVARLAGVSMATSVRVHMAESVAPARDVLOAIGTLTVYEDLSARKKRAQGRKPTLAV ROPLCFMLPESALATGRRGOFEOAMRDRGLAGVFNHATGDHDTLASLNN AHRKSGPDDVILICGNDBAAVAVOVLAKGVRIPQVAVMGEDNLVGVHFLPLPTITQDPHDIISREALHIIEGREGVTRTI				
D70044	ck: 4428 len: 243	1	transcription regulator GntR family homolog		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{124}(P)xx(S)(L)xx(I)(T)(R)xx(L)xxx(I)x{103} MNIRKQSPILITQIMQIMQIKTOIKNGELOKMDPLPSREHVAEGEISMTVRQALSTLVNEGILYR ISRAKQLEPSAATTEBANILGIQKAPVLLIKRTIYLDNGIAFEHASYRGDRY				
S48826	ck: 9647 len: 215	1	NADH dehydrogenase (ubiquinone) (EC 1.6.5.3		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{36}(A)x(T)(P)xx(V)(T)(R)xx(P)xxx(S)x{163} NAMITRTARTRRLPVLOSHRAAASHLHTSLPALSAPATPISITRPPPTSSAPPGLSKTAEEVTL VDIYVPGCPPTAEALTYGLTLOLQKIRNRKDFLHMWNK				
S75086	ck: 4153 len: 342	1	iron-stress chlorophyll-binding protein - S		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{43}(A)x(T)(L)xx(I)(S)(R)xx(P)xxx(M)x{283} MOTYGNVQVEMWAGNARFADOGSLIAHVAQAALITAFMGAFITLIESRFDPTQMGDOGLIIT SOPITLPLVYIGYOTHFASISLIEDVAGIIEFGLIGGIIHILVPLGAKKV LIFGEALISLSGIGLAFVAAVCAVNTLAYPPEFPGPLAKLGIFFYFADTVELPMHATSRAMLANAHFLEAFFLQGLHMH				
A64834	ck: 3224 len: 182	1	probable fimbrial protein-like protein yobC		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{44}(A)x(T)(L)xx(I)(T)(R)xx(V)xxx(P)x{122} MTTKKSYULAFITIVVCAITSSVMAADNATLTDSSVTFNKNVLAFACTIVAAAKDSVTLTPVYSATK HINNY				
D71484	ck: 1871 len: 179	1	probable ribosome releasing factor - Chl		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{68}(A)x(T)(V)xx(I)(S)(R)xx(L)xxx(L)x{95} MTLAAEKEMVGLITFEQKEGRGRTCKAHPALVETVVEYVGTIMRLSDIASISVSDMRLL				
S74877	ck: 7310 len: 328	1	phosphate transport system permease prot		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{151}(L)xx(I)(L)xx(I)(S)(R)xx(I)xxx(S)x{121} PSMLVAGLVLTWITITIASIRILLSVSPSLSSAMALGATRMETISLTPSA SSGIIIGATILALGRALGETMAVTVIGNSITSLAPLITPSVLAQFAVADLHIGAKMLALITVLIGNSLAVIMV				
A46259	ck: 9097 len: 438	1	recombination protein recA homolog - Ara		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{200}(L)xx(T)(A)xx(M)(C)(R)xx(A)xxx(I)x{222} DSQVLSTLKNPSFTPLSPLEPPTPCSSFSLSRSLYSPTVYAAKLSHKISSEF YSKALGVYDENTLVCOPDNGEMALLETADRMCRGAVVLIQVDSVALTPAEIIEG IGMOQGLQARLMSQALRMSGNASKACITLIFLNOIRYRIGVYGNPESGIALIKFPASVRLIEISACKIKSSKDEDIGLR				
J00661	ck: 6434 len: 424	1	impB protein - Salmonella typhimurium pl		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{145}(A)x(S)(A)xx(A)(T)(R)xx(P)xxx(G)x{263} MFALDINSEVASCCEVPEPDNRNPFVVLNNGCVIARISPEAKALIGMGPFVQVRQML GLDPYMGVGRILTEKNAIGINTALQALQANTAFKRNKSTLIERVTEIENGE SCISLEAPPAKQIYCSISFERITDDAHQAVVOVABEAAKILGEROYOTVTPVTSFPAVKPCYSNAAEKLEPLPTQ				
F71946	ck: 9168 len: 343	1	flagellar motor switch protein - Helicob		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{42}(L)xx(S)(I)xx(I)(S)(R)xx(V)xxx(G)x{1285} MATLRFQKQAOBDELSMERKARILILVQGEDTCEILRHLDISITIEISKQIYQJLNGDKOI EISPOVYKRVSTVLEKLESLTSYKIEVGLAVAEVITNRUGQSAKITLARISSV DNKLGAIRKEMTFEDIDAKLNFALREILKVAKDRLSLAKITODLDFKFLNNSSNAACQFVEMOYIGANKINDVYAOAR				
A37763	ck: 2312 len: 147	1	Agrobacterium tumefaciens		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{125}(L)xx(S)(I)xx(V)(S)(R)xx(I)xxx(S)x{6} MSQSRPTSSDIAVNORECVKVEGRVYSTRLSRAEYESFHSQARLLGLSDMAIRVAVRRIG				
A25063	ck: 4145 len: 147	1	hypothetical virdI protein - Agrobacteri		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{125}(L)xx(S)(I)xx(I)(S)(R)xx(I)xxx(S)x{6} MSKRTVTSSETAINQHRSLNVEGFVAVSARLSAEYEFYSQARLLGLSDMAIRVAVRRIG				
S06883	ck: 2761 len: 147	1	Agrobacterium rhizogenes		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{125}(L)xx(S)(I)xx(V)(S)(R)xx(I)xxx(S)x{6} MSQSRPTSSDIAVNORECVKVEGRVYSTRLSRAEYESFHSQARLLGLSDMAIRVAVRRIG				
S36564	ck: 5828 len: 368	1	E2 protein - human papillomavirus type 4		
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{63}(V)xx(P)(P)xx(I)(S)(R)xx(A)xxx(I)x{289} NMQCFKESLSESRISAQDRIIDHYRSDSDINPSTQILRENAITLTPAREHGITKLNHO FKSECEKGNSTWEPYGVGNVADICNDMSKTSDDIVTSQITVRLQHASISTPXT				

ASVGTPEKPIOTPAKPRROCGLTGHHGVNTHVHNPCLCSSTSNKRRKVCSGNTPTIHLKGDKNLSKCLRRLKRYADHSEIS

1 S36576 ck: 4491 len: 368 i E2 protein - human papillomavirus type 52

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(157)(V)x(P)x(V)(S)(K)xx(A)xxx(I)x(310)
x(57)(V)x(P)x(V)(S)(K)xx(A)xxx(I)x(295)
MESIPARINAAOEKLELDYEDSDNLNOJIEHMKRLMECELYEKAELGITIGHQVPPPAVNSK
1: YCIVGVMEVHVGQVIVCPASVSSNEVSTETETAVHCTEKSASVAGAKDTILO
PPQKRRRRDPVDSHNTKYPNNLLGQOSVDSTRGLVATECTINKGRVATHTCTADILHLKGDPNLSLKRYRVTHKSLYVLSSTW

1 S36552 ck: 4819 len: 383 i E2 protein - human papillomavirus type 3

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(57)(V)x(P)x(V)(S)(K)xx(A)xxx(I)x(310)
METLANRLDVCODKLELXEDSDKLENOJIEHMKRLMECELYEKAELGITIGHQVPPPAVNSK
1: RYVGDTGTMVHVGQVIVCPASVSSNEVSTETETAVHCTEKSASVAGAKDTILO
PEQKRRLLEYVGEQOQOQOQOQOHTGPAPOTTERKROFLDTRDRDTTCPHRIGHRSDPCVPYHLGSDNCLKCPRYRL

1 S36535 ck: 273 len: 376 i E2 protein - human papillomavirus type 10

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(57)(V)x(P)x(V)(S)(K)xx(A)xxx(I)x(310)
METLANRLDVCODKLELXEDSDKLENOJIEHMKRLMECELYEKAELGITIGHQVPPPAVNSK
1: CVYGETGKMEVHVGQVIVCPASVSSNEVSTETETAVHCTEKSASVAGAKDTILO
GPEQKRRLLEYVGEQOQOQOQOQOHTGPAPOTTERKROFLDTRDRDTTCPHRIGHRSDPCVPYHLGSDNCLKCPRYRL

1 A44215 ck: 5996 len: 303 i EUS1 protein - equine herpesvirus 1 (strain

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(62)(L)x(T)(L)x(X)(A)xx(X)(C)(R)xx(V)xxx(G)x(1225)
MGVLTAVTVTVDRKALPNSIDVGLMEFLRQCFVLASEPLGIPIVRSADLYRSSLILTE
1: EHNHPLKIRAGSGEDTVVGECEGFSKHSNSLVHPPTNRYAVVDPARLEIPAP
GRPLRRRPSGGRARRRRRRAPARASTAAATPRRDPAPARARAGDVTMMERLIMVFGFTSTR

1 F43674 ck: 501 len: 372 i US7 protein - human herpesvirus 2 (strain H

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(150)(V)x(S)(A)xx(A)(S)(R)xx(L)xxx(L)x(206)
MGRSLQSLALIGLWCVATGLVGRPTVSLVSDLVAGAPGQFVEEDLRFVGLHFGVAGVPH
1: DYGSCDPAQLPFSAPRLGPSSVYTPGASRPTPPSSPRDPTPAPADTETPA
PASEGIAAPPNTRASASRHLTVAAOYIOIAIPASIAVFLQSCICFIHRCORRYRRRGOIYNGGVSCAVNEMAMARLGAELRSH

1 S58346 ck: 2806 len: 256 i coat protein V1 - tomato yellow leaf curl Y

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(20)(L)x(S)(I)xx(X)(S)(R)xx(A)xxx(V)x(1220)
MSKRPADMIISGPVSKYRRLSISPSISPSRAVAVITVGTGKEKMANRPMKRMFMRFGPDVPR
1: VLRRKHAATVNGGQVASEKQOALVKEFVAVNNVYVNOQEAKEYNHSMNLMYMAC
THASNPYATIKIRIIFYDYSTN

1 J01248 ck: 566 len: 106 i hypothetical 12k protein - chrysanthemum vi

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(22)(V)x(S)(I)xx(X)(S)(R)xx(L)xxx(G)x(68)
MPLTPPDHDKVLLVAIGLSTVASTILYTSRNTLPQVGDHSHLPHGQVYKDTGTTIYGGPRK
1: S41285 ck: 3148 len: 355 i coat protein - sweet potato G virus (fragme

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(45)(P)x(P)x(P)x(X)(S)(K)xx(I)xxx(I)x(294)
SABEITGKGTGNGRGRGCTVPPPPPPGAPGTGLDPLPVGCPPPPAASKPPIIEIILQPS
1: NAMEGVRDEGVNDEMGLINGLWMCIEGNSPINMVMGDEVTYPIK
PLDHAAPTFRQIATHSDIAEAVIERKRNKIKAYMPRYGQRNLDTMSIARVAFDYELHSTPYARABAHQMKAALKNQNLFC

1 E71669 ck: 3620 len: 433 i preprotein translocase secy chain (secY

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(197)(I)x(S)(H)xx(L)(S)(R)xx(A)xxx(L)x(1220)
MGONFSKSSSDNLNFTLFTLMLILICRSESTPDPIDSLNSVAEKNDPGLGEMNLSG
1: IONGSTLIIFIGIISGVSAIISMEFELSKRGALEPLATVTCIGVLLIAIIFPE
KAORRLLYOYPRKQVGRKRTIGGATMPKNTSGVPIPFASISLLEPTTLASENSNSDMSNLTYVYLGHRPYLLVYVLI

1 C69632 ck: 4284 len: 258 i transcription repressor glcR - Bacillus

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(24)(I)x(T)(L)xx(V)(S)(R)xx(A)xxx(L)x(218)
MYGEERVAIIDEFLKOHNRITTEOICITLCOVSDTARLDLVKLEQNALIRTRGAILPTVHQ
1: HGITIAHEEDGMYRKMKIOQAKOVIALADHSLKSGTSFYQVAMEIDLLITDRCLP
NQACDLDLNGVELVTEDEGCKD

1 F69260 ck: 6950 len: 244 i nitrate ABC transporter, permease protel

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(71)(A)x(X)(A)xx(L)(S)(R)xx(V)xxx(T)x(157)
MQAKRDKTGLITLIDMSRADIITMTAITMALGLFALIVY
1: ISITATAVLPFMAESFATTSGLDTLIDMSRADIITMTAITMALGLFALIVY
ESAERRVCNML

1 A70001 ck: 6987 len: 253 i ABC transporter (ATP-binding protein) ho

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(52)(L)x(S)(I)xx(V)(S)(H)xx(I)xx(G)x(185)
MYLEANKRKISYGNKINROEVKLDIHIKEGEFVISINGASGCKTTLNLSIDQVSHGT
1: SASDLNKLKLSQLNOKRNATIIIMTHTDHPVAASCGRYIFIDGOMTOLNKGODRO
TFEODIKTQGVIGVQVQHEH

1 A70039 ck: 1714 len: 301 i ABC transporter (ATP-binding protein) ho

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(111)(L)x(S)(L)xx(L)(T)(K)xx(A)xxx(L)x(173)
MKKRAVYVSNVTKRFOKRTVANNISIEGELAAALGNGACKTIVSMILGLKPSGEIK
1: DOGTTITFTSHYIOEADDAORLFTFEGLVADGSPQIKRTOQDSFTLHSS
ESTERLSCHEFEVRIHEHERITTIQTSNIDKVALIPEBHARIDRIEOLADEAFQOLADGNREAM

1 G69260 ck: 9269 len: 243 i nitrate ABC transporter, ATP-binding pro

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(155)(L)x(S)(L)xx(L)(S)(R)xx(M)xxx(L)x(172)
MITVAGISKEFGELKALDGVSEFVEGECALIGPGCKSTLIMAGLKPSEGEVLVDOR
1: TKTMTLVATSHIEBAVFLGRKIVLTERGPRVAVVNDREAGDESRYEEVEFFERC
KLKQIIRAT

1 S56325 ck: 1039 len: 252 i hisP-like nucleotide binding protein phn

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(4)(L)x(S)(V)xx(L)(T)(H)xx(A)xxx(G)x(1232)
MNOBLISVNNLTHHXARGKGFVSVDIPGPGTIVGSGSKTTLQKLSARLTPQOGEH
1: VSVQARLDLRLGLVLEMLNAVITVTHDGLARLADRLVAKQGVESGLTRVY
LDDEHPTQLVLSVLYON

1 S73545 ck: 2944 len: 450 i ATP-dependent RNA helicase dead - Mycopl

<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x(192)(L)x(T)(A)xx(I)(T)(K)xx(I)xxx(L)x(242)
MDSFNEFLASGPAIATLKNNDKNNOPTIOLQAIPOPLHONILVHSPGTGKTAVFGIPVIE
1: FEOJIAVVSAGYNEETIETAKOITRKNIGIFLAABELKONAEPEKNTIDQVCLFEN
RKKOALYLSQTRAKSIIVFCDTKRLVDELCLFKLNKVYKTPPHGKRAQIRERNKLPAANTAPVIVLTDLIGAGIHVEGV

1 S66920 ck: 2958 len: 482 i probable RNA helicase CAs/6 - yeast (Sac

1	1:	MMSNRKQHFAMHPHNPENKYPSSLHSSSEAIRACLPPTPLQSLNFAFSLUDETLLARAEALAAVDI GGGGILGSAHPHPPHSHLGHSHPPHAAAMNPPSGIPHLVAAAHHGAAAAA AAASAGQVAAAAGAAAGAGLASICDSIDPRELEAFERKORRIKLGTVQADVGSALANKIPGVSLSGSTICRFESLTLSH	x{193}(M)x(S)(L)xx(L)(S)(H)xx(A)xxx(M)x(211) MMSNRKQHFAMHPHNPENKYPSSLHSSSEAIRACLPPTPLQSLNFAFSLUDETLLARAEALAAVDI GGGGILGSAHPHPPHSHLGHSHPPHAAAMNPPSGIPHLVAAAHHGAAAAA AAASAGQVAAAAGAAAGAGLASICDSIDPRELEAFERKORRIKLGTVQADVGSALANKIPGVSLSGSTICRFESLTLSH
1	1:	MSGSGGKARAGMAAGAGSSGGRSLSRRTNMSGRKHTTTFSSVSDILSPLEESYKKYMGEGGGGAF DYSKMMAPLPSAPRRKRVLSQNOYVLEPRFPOOKRYSAPREHNLASKYHLTPRT QVKIMFQNHRIKMKHQADKAAQOQLOQDSSGGGGGGGTGCPQOQOQOQSPRAYVPLVYKQKPCQAQAPAPGAASLQGHQOQOQ	1: MSGSGGKARAGMAAGAGSSGGRSLSRRTNMSGRKHTTTFSSVSDILSPLEESYKKYMGEGGGGAF DYSKMMAPLPSAPRRKRVLSQNOYVLEPRFPOOKRYSAPREHNLASKYHLTPRT QVKIMFQNHRIKMKHQADKAAQOQLOQDSSGGGGGGGTGCPQOQOQOQSPRAYVPLVYKQKPCQAQAPAPGAASLQGHQOQOQ
602321	ck: 1141	len: 401	thyroid transcription factor 1 - human
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{112}(L)x(S)(L)xx(V)(S)(R)xx(L)xxx(M)x(1213) x{136}(L)x(T)(L)xx(L)(T)(R)xx(L)xxx(P)x(1264) MLKVLHSDIHGSGSHGHNATGNTLRDLDFOSLRICIDRAIAPADVLFEGGAPFDPATPF IKSLDPOPTLLILAMDRATFGEKTLVSGKGTIPALINRPEPYYALGVHV KKNQNNPHNPIYVPGSIEERDEEKEEDGYIMLEISKRKVMQFCPLPAPRFKTIKVYVTEATPQGBELMALAKTADNAVYRL	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{107}(M)x(S)(L)xx(L)(S)(H)xx(A)xxx(M)x(208) MNSVPCHTVPLAHNNHHHHHQALEDLDHLSPLSLMAGAGARRGAGGGGAGHDAAGGGGF DPRELEAGSEFKORRIKLGTVQADVGSALANKIPGVSLSGSTICRFESLTLSHN NMIALKPILOAMLEAEGPSEKMKRPFELFNGEKKRKTSTIAAEKRSLEAVFAYVOPRSEKIAIAETKLDKKNVYVMEFCNOHQK
578452	ck: 3745	len: 331	POU-domain protein rdc-1 - human
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{107}(M)x(S)(L)xx(L)(S)(H)xx(A)xxx(M)x(208) MNSVPCHTVPLAHNNHHHHHQALEDLDHLSPLSLMAGAGARRGAGGGGAGHDAAGGGGF DPRELEAGSEFKORRIKLGTVQADVGSALANKIPGVSLSGSTICRFESLTLSHN NMIALKPILOAMLEAEGPSEKMKRPFELFNGEKKRKTSTIAAEKRSLEAVFAYVOPRSEKIAIAETKLDKKNVYVMEFCNOHQK	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{107}(M)x(S)(L)xx(L)(S)(H)xx(A)xxx(M)x(208) MNSVPCHTVPLAHNNHHHHHQALEDLDHLSPLSLMAGAGARRGAGGGGAGHDAAGGGGF DPRELEAGSEFKORRIKLGTVQADVGSALANKIPGVSLSGSTICRFESLTLSHN NMIALKPILOAMLEAEGPSEKMKRPFELFNGEKKRKTSTIAAEKRSLEAVFAYVOPRSEKIAIAETKLDKKNVYVMEFCNOHQK
576329	ck: 4655	len: 416	probable phosphoesterase (EC 3.1.-.-) - Syn
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{136}(L)x(T)(L)xx(L)(T)(R)xx(L)xxx(P)x(1264) MLKVLHSDIHGSGSHGHNATGNTLRDLDFOSLRICIDRAIAPADVLFEGGAPFDPATPF IKSLDPOPTLLILAMDRATFGEKTLVSGKGTIPALINRPEPYYALGVHV KKNQNNPHNPIYVPGSIEERDEEKEEDGYIMLEISKRKVMQFCPLPAPRFKTIKVYVTEATPQGBELMALAKTADNAVYRL	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{136}(L)x(T)(L)xx(L)(T)(R)xx(L)xxx(P)x(1264) MLKVLHSDIHGSGSHGHNATGNTLRDLDFOSLRICIDRAIAPADVLFEGGAPFDPATPF IKSLDPOPTLLILAMDRATFGEKTLVSGKGTIPALINRPEPYYALGVHV KKNQNNPHNPIYVPGSIEERDEEKEEDGYIMLEISKRKVMQFCPLPAPRFKTIKVYVTEATPQGBELMALAKTADNAVYRL
771508	ck: 6469	len: 244	probable phosphoesterase (EC 3.1.-.-) CT488
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{99}(P)x(T)(L)xx(L)(S)(R)xx(V)xxx(A)x(123) MRLEFADHLHSLGVEKTEVGEWVGTIHOKEIRHMDIVSDIVCLPDGDISMANREBAOVT THYPVNSDGTGRVSNLLEMDGRVSRCLFGLHKKVPRPFGNIRKIEITLVAA DVVDLPIPOVS	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{99}(P)x(T)(L)xx(L)(S)(R)xx(V)xxx(A)x(123) MRLEFADHLHSLGVEKTEVGEWVGTIHOKEIRHMDIVSDIVCLPDGDISMANREBAOVT THYPVNSDGTGRVSNLLEMDGRVSRCLFGLHKKVPRPFGNIRKIEITLVAA DVVDLPIPOVS
535782	ck: 5295	len: 467	serine/threonine protein kinase - bovine he
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{133}(A)x(P)(A)xx(V)(T)(R)xx(A)xxx(A)x(308) MERAARLRRORRGLMRSRFACVAAEPSSGRLGOSVGAALAPARCAAGSADLYLAANNESPE VEATGPAPAOEHVVLKIGASSTLAEMLRITDHAHVYLRKAVLFHGGELCAVL ARYREDLHTHLMKRIDRPMALPTALQVTRAVLOGIAYLSRIRAHVDYKTEVFLNPGDVCLDGFGAAGHGVTEPRYYGLAGTLETSN	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{133}(A)x(P)(A)xx(V)(T)(R)xx(A)xxx(A)x(308) MERAARLRRORRGLMRSRFACVAAEPSSGRLGOSVGAALAPARCAAGSADLYLAANNESPE VEATGPAPAOEHVVLKIGASSTLAEMLRITDHAHVYLRKAVLFHGGELCAVL ARYREDLHTHLMKRIDRPMALPTALQVTRAVLOGIAYLSRIRAHVDYKTEVFLNPGDVCLDGFGAAGHGVTEPRYYGLAGTLETSN
669230	ck: 3724	len: 257	conserved hypothetical protein MTH973 - Met
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{138}(L)x(S)(V)xx(T)(T)(R)xx(P)xxx(G)x(1103) MLTMMGIVGCAIALILGTFRDGRVPLKGFETFDRLDRERENLASVNDGRAVLDVADMLPEVDI TSVMAATSRILKLKLSLSENIHIGT	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{138}(L)x(S)(V)xx(T)(T)(R)xx(P)xxx(G)x(1103) MLTMMGIVGCAIALILGTFRDGRVPLKGFETFDRLDRERENLASVNDGRAVLDVADMLPEVDI TSVMAATSRILKLKLSLSENIHIGT
440369	ck: 4295	len: 302	trans-acting transcription regulator rbcR -
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{83}(L)x(S)(S)xx(V)(S)(R)xx(A)xxx(A)x(203) MHVSLQOLVFEEVVAHNSYTRAAELHLISOPVAMOVROLEDELGSLFRLGKOVVLTDEAGREV GERAISTLARLAEEFVNRREGSGTROAREFESEGGOTIRHGOMTRNEAVKQAVR SGIGLSVSLHTIELETRRLVTLVDEGFPDRROWYLVYRRKRLSPAAGAFREVLSEAAHRCRLG	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{83}(L)x(S)(S)xx(V)(S)(R)xx(A)xxx(A)x(203) MHVSLQOLVFEEVVAHNSYTRAAELHLISOPVAMOVROLEDELGSLFRLGKOVVLTDEAGREV GERAISTLARLAEEFVNRREGSGTROAREFESEGGOTIRHGOMTRNEAVKQAVR SGIGLSVSLHTIELETRRLVTLVDEGFPDRROWYLVYRRKRLSPAAGAFREVLSEAAHRCRLG
A33960	ck: 1522	len: 306	cytochrome caa3 oxidase ctaA - Bacillus sub
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{51}(P)x(S)(T)xx(W)(S)(H)xx(A)xxx(S)x(1239)	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{51}(P)x(S)(T)xx(W)(S)(H)xx(A)xxx(S)x(1239)
1	1:	MKKALALGVLTTFWMLIVLIGALVTRTSGGGCGGRCMPICRGRFPPELNPASIIEMSHRFA TGAVYRHTESSLACPNVLGSPNLNPTOPHEWONGHRAAALLLVNIVAAVH ATTSYKDOQLTFWGMGIIISCLITITLQALSGIMIVYSELALGFALNHSFFIACLFGLCYFLLLIARFYESROS	1: MKKALALGVLTTFWMLIVLIGALVTRTSGGGCGGRCMPICRGRFPPELNPASIIEMSHRFA TGAVYRHTESSLACPNVLGSPNLNPTOPHEWONGHRAAALLLVNIVAAVH ATTSYKDOQLTFWGMGIIISCLITITLQALSGIMIVYSELALGFALNHSFFIACLFGLCYFLLLIARFYESROS
A45335	ck: 368	len: 297	cytochrome-c oxidase assembly factor cta
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{51}(T)x(T)(F)(L)xx(V)(S)(H)xx(V)xxx(A)x(1229) MKRRLTYVITISIGVILVYLLQALGVTRTSGGGCGGRCMPICRGRFPPELNPASIIEMSHRFA YHRYVLTMTATVSLVYVAQFISGISITVFQMLNLSGLNHTLISLFSALSTMANIITRPSH	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{51}(T)x(T)(F)(L)xx(V)(S)(H)xx(V)xxx(A)x(1229) MKRRLTYVITISIGVILVYLLQALGVTRTSGGGCGGRCMPICRGRFPPELNPASIIEMSHRFA YHRYVLTMTATVSLVYVAQFISGISITVFQMLNLSGLNHTLISLFSALSTMANIITRPSH
G02514	ck: 2327	len: 326	P2 purinoceptor - human
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{177}(L)x(P)(P)xx(A)(T)(H)xx(P)xxx(A)x(133) MDNGTQALGEPPTCYVRENKOLLPPVYSVAVLAAGLPINICVITQICTSRRALTRTAVT LSPPALATHYMPYGMALTVIGFLLPFAALLACVCLACRCRODGPAPVPAQERG KAARAAVVAFAISFLPFHTTKTAYLAVSTGVCVLEAFAAAYKGTBPASANSVLDILFEYFTOKFRRRPHELLOKT	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{177}(L)x(P)(P)xx(A)(T)(H)xx(P)xxx(A)x(133) MDNGTQALGEPPTCYVRENKOLLPPVYSVAVLAAGLPINICVITQICTSRRALTRTAVT LSPPALATHYMPYGMALTVIGFLLPFAALLACVCLACRCRODGPAPVPAQERG KAARAAVVAFAISFLPFHTTKTAYLAVSTGVCVLEAFAAAYKGTBPASANSVLDILFEYFTOKFRRRPHELLOKT
JC4800	ck: 7545	len: 328	P2y6 receptor - human
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{179}(L)x(P)(P)xx(A)(T)(H)xx(P)xxx(A)x(133) MEMDNGTQALGEPPTCYVRENKOLLPPVYSVAVLAAGLPINICVITQICTSRRALTRTAV YDLSPPALATHYMPYGMALTVIGFLLPFAALLACVCLACRCRODGPAPVPAQERG RKAARAAVVAFAISFLPFHTTKTAYLAVSTGVCVLEAFAAAYKGTBPASANSVLDILFEYFTOKFRRRPHELLOKT	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{179}(L)x(P)(P)xx(A)(T)(H)xx(P)xxx(A)x(133) MEMDNGTQALGEPPTCYVRENKOLLPPVYSVAVLAAGLPINICVITQICTSRRALTRTAV YDLSPPALATHYMPYGMALTVIGFLLPFAALLACVCLACRCRODGPAPVPAQERG RKAARAAVVAFAISFLPFHTTKTAYLAVSTGVCVLEAFAAAYKGTBPASANSVLDILFEYFTOKFRRRPHELLOKT
F64471	ck: 3696	len: 415	hypothetical protein homolog M01375 - Me
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{122}(A)x(T)(V)xx(F)(S)(R)xx(L)xxx(L)x(127) MSIKRDSITILHNSNLSKRMALFELFIRFELTAFELGLMPLADTLTIFSSSGIPPAIA YLVGGIFGLYILYKALKRKFEDKRLTIFSNFNDIRISIPALHSS YHFGDINDIVMSIMGFSWISIGYSISRGIFPRASAVSIFLPRISIKTDLISLKGCIIONTIIFSSIFVIGCLFPEIPILI	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{122}(A)x(T)(V)xx(F)(S)(R)xx(L)xxx(L)x(127) MSIKRDSITILHNSNLSKRMALFELFIRFELTAFELGLMPLADTLTIFSSSGIPPAIA YLVGGIFGLYILYKALKRKFEDKRLTIFSNFNDIRISIPALHSS YHFGDINDIVMSIMGFSWISIGYSISRGIFPRASAVSIFLPRISIKTDLISLKGCIIONTIIFSSIFVIGCLFPEIPILI
A47201	ck: 1219	len: 120	bombesin-like peptide - African clawed fr
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{8}(M)x(P)(L)xx(L)(T)(H)xx(L)xxx(T)x(96) MSAVALTRMLPLRFLTHLLLSFTPLPYFCMEFSEDAENIEKIRGNOMAIIGHFMGRKSLD	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{8}(M)x(P)(L)xx(L)(T)(H)xx(L)xxx(T)x(96) MSAVALTRMLPLRFLTHLLLSFTPLPYFCMEFSEDAENIEKIRGNOMAIIGHFMGRKSLD
E65139	ck: 1678	len: 197	hypothetical protein b3434 - Escherichia
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{110}(L)x(T)(L)xx(L)(S)(H)xx(P)xxx(G)x(61) MNEIISAVALLLILNDPILGNLIPFNSVILKHTPEKRRALIMVRELLIALVLMVLFEGEKILTA	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{110}(L)x(T)(L)xx(L)(S)(H)xx(P)xxx(G)x(61) MNEIISAVALLLILNDPILGNLIPFNSVILKHTPEKRRALIMVRELLIALVLMVLFEGEKILTA
IIVMAATQWFLDGIIMNMKG	1:		
548737	ck: 6122	len: 423	kynurenine aminotransferase - rat
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{161}(L)x(P)(A)xx(A)(S)(R)xx(P)xxx(V)x(246) MKRLDARLDDSIDONLWREKTLREVDVNLVGGFEPFSPDPAIQAFOQATSGNFMNLQY LVIANTPNNELGVSRMELELVANLCQHDVVCISDEYQWLVVDGHOVNSISLP GMMDRTLIGSAGKSFSATGKVMGMDPMKHLRTVHONSIFHCPTQAAVAVOCFEREOOHFGOPSSYFLQDPQAMELNRD	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{161}(L)x(P)(A)xx(A)(S)(R)xx(P)xxx(V)x(246) MKRLDARLDDSIDONLWREKTLREVDVNLVGGFEPFSPDPAIQAFOQATSGNFMNLQY LVIANTPNNELGVSRMELELVANLCQHDVVCISDEYQWLVVDGHOVNSISLP GMMDRTLIGSAGKSFSATGKVMGMDPMKHLRTVHONSIFHCPTQAAVAVOCFEREOOHFGOPSSYFLQDPQAMELNRD
JQ2268	ck: 6645	len: 364	O-methyltransferase (EC 2.1.1.-) - maize
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{99}(V)x(T)(L)xx(V)(S)(R)xx(A)xxx(S)x(1249) METSPPNYSDDGSLVDGGAAGAAOATSKAFPHVKSVDLA GLASDQSLIVDAIKOSAEVFGOISIVDAGVAGAAOATSKAFPHVKSVDLA HVAKAPFTHTDVQFIAQGMESIPPAADVLLKSVLHMDHDDCVKILKNCKKALPREAGKGVYIIMVVGAGPDSMKKHEMOAI	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{99}(V)x(T)(L)xx(V)(S)(R)xx(A)xxx(S)x(1249) METSPPNYSDDGSLVDGGAAGAAOATSKAFPHVKSVDLA GLASDQSLIVDAIKOSAEVFGOISIVDAGVAGAAOATSKAFPHVKSVDLA HVAKAPFTHTDVQFIAQGMESIPPAADVLLKSVLHMDHDDCVKILKNCKKALPREAGKGVYIIMVVGAGPDSMKKHEMOAI

1	G70165	ck: 3950	len: 315	1	aldose reductase homolog - Lyme disease sp1
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(150)(M)X(S)(I)XX(V)(C)(K)XX(V)XXX(G)X(149)
					MNNLKNKINTYSLIIGSQGGGFGFQVEKTEAKILKLNKADGIRNIDTARAYNGISEKLTIGE
					1: VIEPCEDNNIATISYSTIAGLLSKANIKRKNRKNRNDIFETLILFKEIMPEYTLKT
					INKEIRIAKININLTIELLYSWLTKRTKLSGFEVSGKENVSNFKAEINDKVEITTSILDNFNHOTKNFPLENKKI
1	G69076	ck: 4836	len: 251	1	molybdopterin biosynthesis protein Moeb hom
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(93)(I)X(T)(V)XX(F)(S)(R)XX(V)XXX(I)X(142)
					1: MPEYEGMAVMEWVSROMGLLSRADQDLKLDSTVIGCGGIGGAAYEMLARMGVSLRTIDSDVE
					PSVGLLETDADYKAKLBDLSKTKTPVPIGPAVNTGCLAAEVLKLTGRGVIAIAPR
					1: MKFDFLQGEPEFRIVELS
1	S66490	ck: 2843	len: 332	1	hypothetical transcription regulator pepa-g
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(110)(I)X(T)(L)XX(V)(T)(K)XX(V)XXX(I)X(306)
					1: MNRHRIEIDITATLACVTTATVTSRYRISPRKVKAKETGERIAKIMEINVIENRAPGMLLAQSVTL
					KVYRHKIIVSGSDPDRDSORVQGYCDAMETNISPIAMNPAISIIHGMQMRD
					1: ALSANPDQVCTNDINDIMGALLLCRENNLAVPQISIAEGHLEIGRQNPISLAVITPRDIGMAAOMLSKIKNNNDNHVTD
1	H69823	ck: 1929	len: 302	1	conserved hypothetical protein yhc2 - Bacil
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(124)(L)X(T)(A)XX(A)(S)(K)XX(I)XXX(M)X(262)
					1: MNQGRLETLIENKQDGVLESLVLTAKSKAPVYIDMNSHQQIKVNHESVLMNMTYKKGDRVFI
					LRTKGTINPPIGRDSHPTRRRVSPGQTAATHEKVAASKEKLELVELELTG
					1: RTHQIRVHLASLHPLTGLSLYGGGSKLLNRQALHANKVQAVHPITDELIVAAEPFADMKNICRTYES
1	A64223	ck: 590	len: 308	1	probable pseudouridine synthase - Mycoplast
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(116)(L)X(S)(L)XX(L)(S)(R)XX(V)XXX(I)X(276)
					1: MKOCFVATTKRDLSDSLASLNSRYKVKLIVDGLVSNKGIKTKNGWLVQPEDRVHVMSELE
					INAPLIARVNNKVMFRIAGTAKAKKQAITKFKVIONNEKALISLELGRTHQIR
					1: HUKFIQHPVNDPLYIKSKSKNSQGFHANRIGCFIHPTLNKPMDFAPLEPFRSFKLSNLISLDPVLVFK
1	S73869	ck: 6533	len: 309	1	hypothetical protein ycec - Mycoplasma pneu
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(118)(L)X(T)(L)XX(L)(S)(R)XX(V)XXX(I)X(277)
					1: MEQIFSTIQRDLDTFATILNLSRYKAKLIVDGLVSNKGIKTKNGWLVQPEDRVHVMSELE
					ISAPLEKRVNKKVMYVGNSSNKAFAKATKFEVLNNEKALIKCELGLGRHOIR
					1: VHQFIKHPVNDPLYIGLXSEQATEYGOYLHQQISFIHPTLNKEMGEAOLDKIFSDKLDNLKANSLYALFO
1	B64026	ck: 2974	len: 164	1	hypothetical protein HI1338 - Haemophilus
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(1106)(V)X(S)(V)XX(I)(S)(H)XX(L)XXX(I)X(42)
					1: MNIFIMHGEAEVANSDKARHLTVYGSKOAFLOGWLMKOHSLTVINSLSRLIVSPYVRAQETFH
1	B65007	ck: 9360	len: 161	1	hypothetical protein b2340 - Escherichia co
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(199)(V)X(S)(V)XX(I)(S)(H)XX(L)XXX(I)X(46)
					1: MOVFIMHGDALDADSDVSRPLITNGCDSRLMAMWLGQVIEIRLVSPFLRAEQTLBEVDDC
1	C64882	ck: 51	len: 299	1	probable transcription regulator ycz2 - Esc
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
					x(1166)(V)X(S)(P)XX(F)(S)(R)XX(V)XXX(V)X(115)
1	E72058	ck: 4148	len: 202	1	probable phosphatase/kinase - Chlamydia
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(30)(V)X(S)(A)XX(I)(S)(H)XX(L)XXX(I)X(156)
					1: MKLKLKSTIGDSSGKTEACQVDELGAIVASADISHSFLIPATRIKGRVIDLSDVYVD
					1: VENNGTKRELHQKIEEYFALKAL
1	B71509	ck: 4232	len: 202	1	probable phosphatase/kinase - Chlamydia
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(30)(V)X(S)(A)XX(V)(S)(H)XX(L)XXX(I)X(156)
					1: MLDLKLKSTIGDSSGKTEACQVDELGAIVASADISHSFLIPATRIKGRVIDLSDVYVD
					1: IENNGTKREERKVKOCFKALNGKI
1	D64151	ck: 1410	len: 430	1	hypothetical protein HI0404 - Haemophilu
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(1157)(L)X(P)(L)XX(F)(T)(R)XX(L)XXX(A)X(1257)
					1: MDLSIEDIKLOKATQAFIALISGLDSTVLISLFAKLQCKQPHLPILSIRAIHIGLSPN
					1: LNWITDSNEDNRYDRNFLRNLPELBRMWHFIDAVQSRASQHOFEQOOLINDL
					SEITEHCKQKQKCOFQYSLAKQTLNRWMLAENQLEMPKSRQITQIINDVIFAKEANQFQLVNKKVIRRQDSLYLTKP
1	S75967	ck: 3587	len: 271	1	hypothetical protein - Synechocystis sp.
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(200)(L)X(P)(L)XX(L)(S)(K)XX(I)XXX(A)X(155)
					1: MEKDDKDFRGVGBEOKTLPCTGPGSSNDRHLNIEKSMITAVVLLSGGLDSTVAAIAKR
					1: LATTOGLAUSKSKVGEPRIOGLAPLIEISKLELGEVPIAGTWSQIAGG
					EEPCGRDSCRDLQALIEAGHPELASAGRLWREKVD
1	D69014	ck: 6260	len: 222	1	conserved hypothetical protein MTH1108 -
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(1159)(V)X(P)(V)XX(M)(T)(K)XX(I)XXX(G)X(47)
					1: MRAISLISGDSAVATALMDEVEIHAITFYDGORSAMELEVARRSEHIGIEHTLLDQW
					1: VGLPELTYSCYAGDRVHCVCSCMRRAFEIAGIDPTEYRE
1	C65012	ck: 4986	len: 285	1	hypothetical protein b2382 - Escherichia
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(1156)(V)X(S)(L)XX(L)(C)(R)XX(V)XXX(P)X(113)
					1: MKAGTLPADQGFEDTSGVINDOLGRLVWFVFSQPSASLPVGSICIDFPRLDIVLRGEYGLL
					1: RQDFIYHSCWYODNYAQLTRSVNAQFNTINPHLSKLPFHOGCMRTIEYVRV
					1: RAAKARMLQKTHLSIHEVAKRCGFPSDIFCVFPROGLIPGETSARFQ
1	A70606	ck: 144	len: 358	1	probable electron transfer protein Rv357
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(1138)(I)X(P)(I)XX(I)(C)(K)XX(L)XXX(G)X(204)
					1: MTEAIGDEPLLDHLEIQIAEIVDEDEASLVFAVPSDDPEIPRRLRYAPQGLITLNP
					1: IAAKYPDLRLTVLHMLLESLOGIPESASALAKVATYTPFICGPFQMAAADALA
					1: ALKVPAQOVHTEVEKSLSDPFAAVKVDSDGAPATAVELDQGTHTVSMRTAKLLDVLAAAGLAPFSRGHCGACACTLR
1	C65112	ck: 597	len: 220	1	sigma cross-reacting protein 27A - Esche
					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
					x(125)(V)X(T)(L)XX(I)(S)(R)XX(A)XXX(C)X(179)
					1: MITMKRIGVILSGGVGDSGSHHAAVLTLLAISRSQAQVAFAPKQGVYINHLTGRAMET
					1: VQDIYVDDNKNIVTTPAVMLAQNIAEASIDKLVSRLVLA

1	JC5603	ck: 1627	len: 270	1	ESI protein - zebra fish	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(56)(A)x(T)(M)x(L)(S)(R)xx(A)xxx(I)x(198) MLASRLALAOAAMLVORACALMHGDMGDMNGNTINIAVFGSCGMDGTDIHEAATMYHLNRN 1: LACRYLPSELEVTEDESSRGMGRWPTNTNVOAKMSGARNHNTREPEAYVDEK NKVISTPTMETEDYHYHIFDIGNMVKHVMKMTAK
1	T09924	ck: 1428	len: 293	1	cytidine deaminase (EC 3.5.4.5) CDA6 - Arab	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(80)(L)x(S)(M)xx(L)(T)(H)xx(V)xxx(G)x(197) MKPYTPSEAEERGVGRSDLPRLDKAMSLARAPSTFRVGVAVGLTSSGEVFLGVNVEPRLPH 1: ECSHTRCRAIAAANKSVAPEKSCGSGVAVGKMYEVAYNPSLCPVRA ALYDFVARGGKFEITEVLVEKMDVVSQEPATRTLDKIAKCDKVLHCYKTKN
1	S55551	ck: 2728	len: 257	1	cytokine-inducible protein C1S - mouse	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(179)(P)x(P)(A)xx(M)(S)(K)xx(A)xxx(S)x(62) MVLCOGSCPLAVEOIGRRPLMAOSLELPGAPMOPLPGAFPEEYTEETPVOAENEPKVLDEPD 1: PAPTPALPKSKODAPSDVLPPIPVATAVHLKLVOPFRSSARSLSOHLCLRYNL VADVDCPLPRRWADLYLRQYFQFL
1	JC5626	ck: 1738	len: 198	1	STAT induced STAT inhibitor 2 - human	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1159)(A)x(S)(L)xx(L)(C)(R)xx(I)xxx(T)x(123) MTLRLCEPSNGGGGTGTSOMGTAGSAEPEPQARLAKALRELIGTGWTGWSMTVNAEKLEKAEH 1: IMGLPLPTRLKDYLEBKQV
1	JC5760	ck: 1381	len: 198	1	cytokine-inducible SH2 protein 2 - human	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1159)(A)x(S)(L)xx(L)(C)(R)xx(I)xxx(T)x(123) MTLRLCEPSNGGGGTGTSOMGTAGSAEPEPQARLAKALRELIGTGWTGWSMTVNAEKLEKAEH 1: IMGLPLPTRLKDYLEBKQV
1	A41900	ck: 4304	len: 299	1	cyn operon regulatory protein cynR - Escher	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1192)(L)x(S)(A)xx(A)(T)(R)xx(I)xxx(C)x(91) MLSRHINVFALVHSGFTTRASALHVSQALSOQITROLESIGVPLFRSGRTIRLTDAGEVWRG 1: HEQVALSRHDEKLVLSAEFATREQIDHYCEKAGIHPOVTEANSISAVLELIR TSISTILLPAIATOHGDKAISIAPLLERTAVLLRRKNSMOTAAKAFLLHALDKCAVVGNESE
1	F70721	ck: 7074	len: 164	1	hypothetical protein RV1829 - Mycobacterium	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(49)(V)x(P)(P)xx(L)(T)(H)xx(I)xxx(I)x(99) MGEYRVGIRVEOPONOPVLLLEKANDRILPIWIGOSEAAALALDOQVEPRPLTHLINDLIR
1	S56618	ck: 3873	len: 173	1	yjiX protein - Escherichia coli	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1144)(V)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x(13) MLIMHVVCAITNPAAKIOALIOAFHEIFEGSCHIASVAVESGPEOPGSEETRAGANRVANAR
1	S45255	ck: 4167	len: 54	1	yjiX protein homolog - Enterobacter aerogen	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(224)(V)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x(14) FALGTVNSOHTGIDIGKREGAIGVFTKGLTRSSVYHQAAYVIALSLFHNATIR
1	A70738	ck: 8373	len: 158	1	probable rimi protein - Mycobacterium tu	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(6)(P)x(T)(I)xx(L)(T)(R)xx(A)xxx(A)x(136) MTADTEPVTIGALIRADAOACAELEOLFVGDDPMPPEAFNRRLASPHNHVYGARSGTLVGY
1	F72316	ck: 5138	len: 309	1	hypothetical protein TW0917 - Thermotoga	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(79)(A)x(T)(V)xx(M)(T)(K)xx(I)xxx(T)x(214) MEVLLDSEFLGEMGANDANAEFGVSGGLPYRKATIVASIFVILSYVGAGAGQNISS 1: TSLAVLSQAMVYWFISLIGIHVSSQALVAVLGAGYKGNLGNKYLILSGWLTFRVSGTSLFLLSLIK
1	B70451	ck: 4885	len: 311	1	prephenate dehydrogenase - Aquifex aeol	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(P)x(P)(P)xx(F)(C)(K)xx(I)xxx(L)x(126) MALISSFNPSPPGFECKNIILKLSLSONVLIVGVGGSGFSAKSLRSGFKGKITYDI 1: TDKRLKLVKRVNEDVGVVEYSPHLHIVGVSHLPAVAFALVDLTHMSRP EVDLEKYPGGGFKFDIRIAKSDPIIMWRDIFLENKENVMAIEGFEKSLNHLKELVRAEELVEYLKIKMEID
1	G39845	ck: 2658	len: 256	1	dihydroorotate dehydrogenase (electron t	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(118)(V)x(P)(L)xx(L)(S)(K)xx(I)xxx(V)x(122) MKRAYLVNCSNOOIAADVPOMYKGEVFOFTPGQFHLKVEAVTPLLRRPISIAVDNFEK 1: VIRKKRLEPDIILSCGTPMLKRLKOEYANRVEYLSMEBRMCGIGACFACVCHN ESERSYKVCOLDGPVFAKEVAL
1	J01382	ck: 6785	len: 303	1	hypothetical 34k protein - pea enation m	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1170)(V)x(T)(L)xx(L)(T)(H)xx(F)xxx(L)x(117) MHQIEQDLPDIYHRCASSTFLASLDGLGKASRELSGVLAITSYLVLASIALCAIIPGS 1: THODPRLYLARRNDLSCGEPHVRPARGVLESALDLSSTSKRQIPPLGVSV LHVLGLAERDCILFDIDSNTSYTHRVLEQDNRNADSLFSDILEYHDLALGYSDEDDDLDNF
1	S56961	ck: 8234	len: 196	1	probable membrane protein YJL178c - yeas	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1167)(L)x(S)(L)xx(P)(C)(K)xx(V)xxx(C)x(100) MLCGLTDLVILPGKDAITTOIIDDKNIGFVEETESALTLTKGATMGANSFPAKLEFOCNDN 1: VVSRIIGRSTAOGRGYSAY
1	J01355	ck: 9413	len: 134	1	V0 protein - Miscanthus streak virus	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(118)(A)x(P)(P)xx(I)(S)(R)xx(A)xxx(C)x(100) MGCYFIMGRGSEITNITAPPPMIPISRLASVCACGPRSLPPLADRDQAAKWRHLSVRA
1	J00150	ck: 6657	len: 122	1	hypothetical 13k protein - Pseudomonas a	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(98)(P)x(S)(A)xx(A)(S)(R)xx(A)xxx(I)x(8) MALASPCSRCCCAASASAPATACCPYGASRTCARISIPWIAWRWITCSTATVACARSRR
1	B71217	ck: 8463	len: 108	1	hypothetical protein PH2002 - Pyrococcus	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(125)(L)x(T)(L)xx(V)(T)(R)xx(V)xxx(L)x(67) MKRHLLEKSLIRPLKIDIFVHCHLDTLGCVTRLGPGFSLHLHFHDLPLVKGHSRPSLV
1	T05820	ck: 4391	len: 217	1	hypothetical protein TSK18.150 - Arabido	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(125)(L)x(T)(L)xx(V)(T)(R)xx(V)xxx(L)x(67) MKRHLLEKSLIRPLKIDIFVHCHLDTLGCVTRLGPGFSLHLHFHDLPLVKGHSRPSLV

1	D71926	ck: 2840	len: 142	1	cag island protein - Helicobacter pylori (S
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(42)(L)x(S)(V)xx(V)(S)(R)xx(L)xx(I)x(84)
1					MKTNFKIKILFAMCLITIGMENAPLNADONTIDKIDISPEDMALNSVGLVRDQIKIEIPKRELEOK
1	F71925	ck: 6963	len: 114	1	cag island protein - Helicobacter pylori (S
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(80)(L)x(S)(L)xx(T)(S)(R)xx(L)xx(A)x(18)
1					MKRPISKLNQNFLOFRHSNKHLDKSLYRLFNLSIVIGFLVALFSYGAGVILVPIELFA
1	H64586	ck: 7347	len: 114	1	cag pathogenicity island protein cag15 - He
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(80)(L)x(S)(L)xx(T)(S)(R)xx(L)xx(A)x(18)
1					MKRPISKLNQNFLOFRHSNKHLDKSLYRLFNLSIVIGFLVALFSYGAGVILVPIELFA
1	F64587	ck: 2551	len: 142	1	cag pathogenicity island protein cag21 - He
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(42)(L)x(S)(V)xx(V)(S)(R)xx(L)xx(A)x(84)
1					MKTNFKIKILFAMCLITIGMENAPLNADONTIDKIDISPEDMALNSVGLVRDQIKIEIPKRELEOK
1	A64649	ck: 6457	len: 131	1	hypothetical protein Hp1033 - Helicobacter
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(122)(L)x(S)(V)xx(L)(S)(R)xx(L)xx(V)x(93)
1					MYSKILATSFSLFNSMKRKSLSVPFLSRQISKSVQAKIKIDANNNAIKYPPSVDNVTHNMLG
1	D64710	ck: 2117	len: 115	1	hypothetical protein Hp1524 - Helicobacter
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(113)(V)x(S)(L)xx(A)(C)(R)xx(P)xxx(S)x(86)
1					MKSKITHFTAFISFVLSLFSACKDEPKRSSQSHONMTKITKNPINOANNDIRKIEHEDEKAT
1	E71938	ck: 3806	len: 91	1	hypothetical protein jhp0391 - Helicobacter
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(122)(L)x(S)(V)xx(L)(S)(R)xx(L)xx(V)x(53)
1					MYFKILATSFSLFNSMKRKSLSVPFLSRQISKSVQAKIKIDANNNAIKYPPSVDNVTHNRM
1	A71809	ck: 3861	len: 115	1	hypothetical protein jhp1413 - Helicobacter
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(113)(V)x(S)(V)xx(L)(S)(R)xx(P)xxx(S)x(86)
1					MKSKITHFTAFISFVLSLFSACKDEPKRSSQSHONMTKITKNPINOANNDIRKIEHEDEKAT
1	C71912	ck: 428	len: 307	1	probable outer membrane protein - Helicobac
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(187)(A)x(T)(L)xx(A)(S)(R)xx(L)xx(S)x(104)
1					MKRVLLTLTSLFSLFWHAENGFVLCINFAEGSTIOGOGSIGEKAAENALNQALNNAKNSLFPN
1					YKDMTGRTLDADTLKRAASHIRKSSGLVGMELGASTWFAASNNLPENQVKSRTI
1					FOLQGRGVAFSSDEYDIDRYGDNELGSSVELGVKPAFKVNYSDSDYDKLDYKRVVSYVLTNTYFNKRNK
1	PH0856	ck: 7474	len: 277	1	mauf protein - Paracoccus denitrificans
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(123)(P)x(S)(P)xx(T)(R)xx(V)xx(A)x(238)
1					MVSEYDHGELGAGASVDCCLFPOSASARIRIYLAALAGAAAGVALASAAGPOPLAVALVGA
1					KWVIGGLYGLGIDLTIVYQPLTLYMTLAAFTGNTAHAIYVALFNLGRLLPV
1					AVNALPIPDYRQVQWMLARHOENNALADGAILTLAGGFYVLA
1	I39703	ck: 4597	len: 102	1	tram protein - Agrobacterium tumefaciens
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(13)(L)x(P)(L)xx(L)(T)(R)xx(P)xxx(L)x(73)
1					MELEDANVTKKVLERLPGLIGTRGLPPTDELTITIDAITRHRRLVKEADELFGALPETYKTS
1	JC2568	ck: 7036	len: 118	1	mray protein - Rhizobium meliloti (fragm
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(48)(L)x(P)(L)xx(T)(T)(R)xx(L)xx(L)x(54)
1					LOTHFVPGTGLAVIIGAVIIGGGLFMTFNAPRAIFMGTGSLALGASAPLAVAATKHEIV
1	S27344	ck: 1548	len: 370	1	hupk protein - Rhizobium leguminosarum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(125)(A)x(S)(L)xx(I)(S)(H)xx(A)xxx(S)x(229)
1					MTEFLGAGTIGIDVTSFALACSAVAKNRRRGTLRMFVGRQPEAPVLAQVESLIGFAGSV
1					LADVEDDHVAFGRPRDLTISDAEVYARLDEGVSLSHLSGRILFETGAVRSA
1					SAGLPEAPHLARLRARIGDVRISLTQTLARTGDFCDASCSGTPGAGGVACARGRLVQIETIGDGRUAAVRIAPT
1	S28677	ck: 6944	len: 243	1	hypothetical protein 4 - Rhizobium sp. (
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(93)(L)x(T)(V)xx(I)(T)(H)xx(L)xxx(S)x(134)
1					MSLOLVTGHSRVAAVPEFAAEKODLSLFDQAPVERPTPAAVAFWEGDQARIHFEVVEGTLR
1					TRLDVADYLGMTIETVSNRIKLAGSGVIAIVGRHAIAIIMKALLALADGECDDG
1	AQ8RAHYKA				
1	S26139	ck: 4831	len: 187	1	signalling protein ampd - Citrobacter fr
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(164)(P)x(P)(A)xx(W)(S)(R)xx(A)xxx(T)x(17)
1					MLLENGWLVDAHVPSPHNDCRPEDEKFTLLVHNHISLPDGFSGPWIDALFTGTIPDHAHF
1	LTTSSDKKIT				
1	A48901	ck: 4831	len: 187	1	signalling protein ampd - Enterobacter c
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(164)(P)x(P)(A)xx(W)(S)(R)xx(A)xxx(T)x(17)
1					MLLENGWLVDAHVPSPHNDCRPEDEKFTLLVHNHISLPDGFSGPWIDALFTGTIPDHAHF
1	LTTSSDKKIT				
1	S40867	ck: 3905	len: 248	1	ferredoxin--NADP+ reductase (EC 1.18.1.2
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(129)(P)x(T)(A)xx(P)(T)(R)xx(L)xxx(T)x(203)
1					MADWTGTVTVQVNTQNTSLFSLYTHAPLPPTACQFNLGLGIDIGERNQAVSVNSPDNDL
1					AGSLTGRPATIEGSELESTIGLPMKNETSHVMICGPNQVNRDQQLKTRTQMTK
1					HLRRRPGMTAEHW
1	A64750	ck: 8141	len: 79	1	hypothetical protein b0249 - Escherichia
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
1					x(5)(L)x(P)(P)x(T)(M)x(F)(S)(R)xx(P)xxx(T)x(257)
1					MOHNYSRWMITLAISFSGVSPDLAIVRYIYQIPMAKFMGFSNTEIGLIMSTGIAIILY
1					SVVYLLIGLGFVSDNNNRNSANNEKROFSDILAVLRISTWYSCMVIFGV
1					FTTAAISTNTVILEYMGSLVAASMGVIVINKIFRALCGPLGIGITTYSKVKSPTRVLIQILSVGLLTLTALVTNSNPQSA

1	B65068	ck: 6883	len: 145	1	hypothetical protein b2849 - Escherichia coli
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{99}(A)x(S)(I)xx(L)xxx(L)x{30}			
1:		MRYTMDIESQIHEVMYHMDIVNSDKKKRPILPKELNEMVLTQTSWTLNSRYVNSVKNV			
1	F64963	ck: 7029	len: 359	1	nicotinate-nucleotide-dimethylbenzimidazole
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{137}(I)X(S)(A)xx(L)(S)(R)xx(A)xxx(L)x{206}			
		MOILALMTALIDTSALDSRQRIHIDCLPKVSGSLGELVLAIQLAGMPGLNGIPVGRKAVLV			
		NANTPAAIYSTIGTDPPEVVGALPTDRIKALIDVYRAITLTONPODGV			
		DYLAIVGDFDLVINGVNLGAASGLPTVLIDGFLSYAALACQSPAIRIPYILPHLSSEKARALSHLGLIEPLYNEMHLSGSG			
1:		T00211			
		ck: 5937			
		len: 291			
		1			
		Type II secretion pathway related protein e			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{51}(A)x(S)(V)xx(L)(S)(R)xx(A)xxx(A)x{224}			
		MFEFSRGRGLFKIDVLKRLTPNRLCYLLAGTOLVSIHHPMTQASVPGLSKVSAPET			
		EKKNAIYDNNLTGVITQDNFVRSVDFDIFSFAVYVNTLSGTRILPKHSHSL			
		YNAAGLHMDLAVSVNGSELRTROAQIMKQIPELKEIKITVERDGLYDAFLAVNGEN			
1:		H64726			
		ck: 98			
		len: 216			
		1			
		yabp protein - Escherichia coli			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{8}(P)x(T)(L)xx(M)(S)(K)xx(L)xxx(V)x{192}			
		MKVSVGMVTLNLSKRNIDYKAVSGDKMDVMNIFORLMETLRLHFWMSDKQTEAVYKILFENYNC			
		SNDILYERPGMANNIGVLPRTVLPRTVLTFTVTLVTVLP			
1:		A64862			
		ck: 1943			
		len: 78			
		1			
		ycg2 protein - Escherichia coli			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{7}(L)X(S)(A)xx(I)(T)(P)xx(A)xxx(L)x{55}			
		MHQNSVTLDSAGALTTRFRANKLHTQETIGELVETLINDGRNLSKSKSCATLRLHMAAGEE			
1:		A64882			
		ck: 6021			
		len: 262			
		1			
		yccI protein - Escherichia coli			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{16}(P)x(T)(I)xx(V)(T)(R)xx(A)xxx(A)x{230}			
		MIRKICTTLPVLPPEPITVTRPRARGAFPPGTEHIGKSLGAPLWFPAPASRESGLILA			
		DPLACIEDRHSSELGEMIAQAFELPLVSVGETPGSGWCADLNLHCTAEPF			
		ISSDEASEKYLFPAMANTLFRHPRKDAIRPS			
1:		S20452			
		ck: 9681			
		len: 271			
		1			
		hypothetical protein X - Klebsiella pneumoniae			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{136}(A)x(S)(P)xx(A)(T)(H)xx(A)xxx(C)x{117}			
		SVPGDPIIDMOOLATIKOILASAGRLCLCSADIEREDKVLAMAVHIEGAGFGDEGRDLC			
		NFGNALRADGRSDPTPLTTIVRHIDYIINIGDHALVAGSDFGDITLPELIDGV			
		AGIPLRINTLRASGQDVLVDKLIMRWMLRVLKNVMOO			
1:		S01838			
		ck: 7149			
		len: 220			
		1			
		nifv protein - Klebsiella pneumoniae			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{60}(V)X(S)(P)xx(W)(S)(R)xx(A)xxx(G)x{144}			
		MSDDITLWMLRFLFQSLPDQDQVLDVLAQDSGETLIPERLALITLQPOLAASPSATAVMSPAH			
		MKAQPTTIOAOCEALINTLLAGRLPWLAKRLNRDPLEERVE			
1:		S22619			
		ck: 7734			
		len: 336			
		1			
		hypothetical protein - Salmonella choleraes			
		<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C			
		x{115}(I)x(T)(V)xx(F)(S)(R)xx(L)xxx(S)x{205}			
		MNNKVLMDISWSKKGIGRFTDEISKEELVIRKCAAPLAPGLAVNIFLRKKTVDVFLP			
		VISAFAKADIDPSIKIVLTGPNCDLEKILIOHGISERKVFGEVSEKDDLSLYKG			

1	SLGLVPSLYEGGLEPVEGMACGIPVLTSLTSSLEPVAGDAALVDPDLSADATLKGISLINDSELKRLIOGILRAKRFNMO
1	A53302
ck: 5663	len: 289
1	1
1	heme 5'-region hypothetical protein - Sa
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{22}(A)x(T)(L)xx(A)xxx(A)x{251}
	MSTRLKLTLSRRQVIRKTLALTLISGSHAVAAKEETLKISNGSKRKTGSKRLVMDP
	ATDRDHLQGVLEFDVQDTIKNSLTGSHIIRKIPKIRHSHRSTEQAAEVNKS
	PSIPSVLSTFSTINPEERLTGTFAROKIATIANGLIISYFHWFDQKAKTKR
1:	
S23906	ck: 4054
len: 135	1
1	1
1	hypothetical protein 1 - Shigella flexne
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{129}(P)x(T)(A)xx(F)(T)(R)xx(L)xxx(G)x{90}
	MADVTKGTIVKQWMDALFSLTVHAPVLPFTAGOPTKLGIEDERVOAVSYVSPDNDL
1:	
T14663	ck: 2204
len: 99	1
1	1
1	histone H5 like protein - Yersinia pestis
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{47}(A)X(P)(V)xx(F)(T)(R)xx(L)xxx(A)x{36}
	MKATIVGANSVPSAAYSRRNDLRADDIYERADIVPMKRGTAEVBAKRVQSTEELES
1:	
T14952	ck: 2530
len: 144	1
1	1
1	hypothetical protein - Yersinia pestis p
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{92}(A)x(P)(V)xx(F)(T)(R)xx(L)xxx(A)x{36}
	MKPAKIRLEPQFLGVLGIGIOFVDGISVALLERFIIDORICASMRATTVGKAVSPSAIS
1:	
S70883	ck: 2492
len: 227	1
1	1
1	hypothetical protein 3 - Vibrio cholerae
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{58}(V)x(T)(I)xx(F)(T)(R)xx(L)xxx(I)x{153}
	MARLFEMNLSRVKRRKISLVDAFEIIFESYSAVWVANGVEIHSDSIPYLATVIVTIL
	KTLENTVIMGVQVHDSRAAYLVKDKVDTQIILAVPSASRAALSDCAKT
1:	
S54442	ck: 6509
len: 150	1
1	1
1	membrane associated protein 17.9K - Vibrio
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{35}(L)X(T)(A)xx(M)(T)(R)xx(L)xxx(A)x{99}
	MKIGERARFVLMVVLVVLITANFVSWRBMFDMHTALEMLRQLIDIRAVTYQEWVLOGR
1:	
G64102	ck: 8973
len: 288	1
1	1
1	phosphatidate cytidyltransferase homol
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{108}(P)x(S)(A)xx(W)(S)(K)xx(L)xxx(L)x{164}
	MKQVLSAIVLAAVLCALFEPFALALAGAVAILGIEWTOFARLQKPLIRFTVTFIG
	GKRKLAPVSPGKSMGVIGLITVALAFIHFSSNLTVDGRNTIFGILISVAT
	VAISVGLTESMFRRESGVDSOLIPGHGVLDIRIDSLTAVAPPEFSYFEFVL
1:	
H64157	ck: 4
len: 311	1
1	1
1	sufl protein homolog H10733 - Haemophilus
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{22}(P)x(P)(L)xx(S)(R)xx(L)xxx(P)x{273}
	MPLRSRROLKTAALSTLSTVPPALAAAREKLVVPLEVARGRPVLTVMQETVPLDGS
	GVDDIPLIOMDEFNPDLOLFKRONHGFNRLVAGIAPYLDVARCMIRRL
	NMSLARADLRINDQERLLAODLGLPRAKSVSVLSPGRALIVNMKLTLYLSLAEVNAVCKRKNKYVLLR
1:	
H71692	ck: 1607
len: 405	1
1	1
1	hypothetical protein R358 - Rickettsia
	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
	x{92}(I)X(S)(L)xx(F)(T)(R)xx(V)xxx(L)x{297}
	MOYLSLTLFSLGAGTGLSIAVAFELSLITIGFISFOKEFNKKFIRKILYPRFGD
	ILFSGHINSLTILVLYLSTSDSLASLFGSISGILITILARAKITIFELVLI
	ISLITGSLFVPVIAKQIDOPNLSKRYLAWHFAVANKITIRIPVGFASSKITIFGDNAADYDROEKIHLPLHP

E72060	ck: 2665 len: 252 cct470 hypothetical protein - Chlamydia pneu
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{147}(L)x(T)(A)x(F)(C)(R)xx(L)x(x)(A)x(189) MOICVTGVLRSRPGLKRNHTLTFTTLPGLTFPPFAPKOSQITQCQYRETIYPISGKTTLHRNSRI KOAISEKEEEOIIOALIHAKOFSELIAAEFLPAINEKITIFLDTSOEERKSERN 1: SSEDPIHELRLSKVHPHY	
F72031	ck: 6017 len: 493 cct578 hypothetical protein - Chlamydia pneu
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{189}(L)x(T)(L)x(A)(T)(R)xx(L)xxx(A)x(288) MSTSSSGDGNKNIMSDLVSTPFGVGRPOOKELSNENKIQOIRQRCKNTMESDATIAGASGRK EVIEIGLALAKAQIQLGRTSALSNSISTGASTQDLLEKOAKIKIREXY QEMKAEEKSKLEESTMDTIVTMVAVALTVISAAIFTGCAGLAGLAAGAAGAAAATVTAQTIVAVQAQVC	
S61492	ck: 5418 len: 270 hypothetical protein 2 - Chlamydia psittaci
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{116}(L)x(T)(P)xx(F)(S)(H)xx(A)xx(L)x(138) MSYSPKHNHSKRTERSRNWSMEDIAADYHIQEGEHYKHKEVILPKPLPLMTNLQSEDSLVDIG KLULSRKIDRYLSITVASIVAHPCFKOSESSISFHPSYWTQALSKYGVINEME EWISPRKSIGAHAKEMLNCREEFFLFMIISCITKHN	
JC5205	ck: 8587 len: 160 sulfur-rich protein - Chlamydia psittaci
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{33}(V)x(S)(V)xx(W)(C)(K)xx(I)Lxxx(I)xx(I)111) MLTGENSEGVIDLTKPGLDDVMKENETVOYVLVNSVLMCKAHIVDPITKTSKIYOSRARQITMYV	
A71529	ck: 4189 len: 303 hypothetical protein CF324 - Chlamydia trach
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{81}(A)x(S)(L)xx(I)(C)(H)xx(V)xxx(L)x{206) MKRAHPLUSSISGALPHISSGGVHDSIGDLHSPLSLTPPALSVLTPPEPAOSKSRVQAVANTIG LGHOTSSUAYGYRRKSGPYTGAODEPFDERKERTIRAOIMOCRLTRNMLOIORIPD TLQKRAHHRQGQSOTITTHPLNRISLPEBMEHELKKKPQVYIAMDNKTAQLNQSIVQLYDRPTVIED	
C71569	ck: 1406 len: 300 probable ribonuclease HII - Chlamydia trach
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{167}(L)x(S)(L)x(W)(T)(H)xx(I)xxx(L)x{117) MPSSEFLSLFELSLELBEOUKRGFTSITSPHVPFOGRSVTCVYOSGKTIYOGKGTOEFVEH TTIIDNLAPHGAVFAIDGFASSERVAOLNAVRKKCSDELIOHRAEQVVAAAA STLARDEAFUSSTHLESQYOTIRLLKNGASGVAKYORAKKEILHNKCQVLEKCKTHEFTENEVLGSNQ	
B53203	ck: 1180 len: 43 hypothetical protein 2 - Desulfotribrio vulg
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{23}(A)x(P)(A)x(a)(S)(R)xx(P)xxx(P)x{4) MTVGLANPGANPPDRRHARARIALDPAGASRSRLPDPMGRMAA	
S17812	ck: 2127 len: 192 hypothetical protein 192 - Rhodobacter caps
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{142}(V)x(T)(L)x(L)(T)(R)xx(M)xxx(I)x{34) MOLDEFERRAAHVPEMATHYIPEPAANOIGAADSDRIGRACQYTINISRLQELHLALTQIVTADS 1: WTNDLSLVISEFSLV	
S22631	ck: 1875 len: 166 petp protein - Rhodobacter capsulatus
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{68}(L)x(S)(V)xx(V)(I)(K)xx(L)xxx(L)x{182)	
MAD7CAPGETLFLTEDEQJRKGIEAMFVFYRGTFTADPRILLDQHDYGAAHHRAIHFHINRED	1:
B56883	ck: 8138 len: 213 transcription negative regulator ChrR -
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{122}(A)x(S)(L)xx(A)(T)(H)xx(I)Lxxx(C)x{175) MTIHHUSDALITRAYAAGTISEAFSLVVAIHLSCDECAHAGDALDAVGSLMETAYALASE 1: TPVAREGLDCICLAATAPARLRFSFPKLYOPFRFI	
S36981	ck: 4069 len: 201 hypothetical protein 7 (atpc 3' region)
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{117}(P)x(S)(I)xx(Y)(T)(R)xx(A)xxx(p)x{68) MYLALOPEGEYOITLAPCGSLRPLPAIEBPALGOVPAOHORODPIVMPNSVLANLMARV 1: AANVTFRPLKIMGOMDVDMYORI	
S74634	ck: 2111 len: 393 heme binding protein precursor ycf10 - S
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{180}(P)x(T)(V)xx(L)(T)(K)xx(L)xxx(L)x{197) MLPBRKSGPNHREHMLVQOAILQAALAESEITEKYTFPDGKRAPDSQGVNTYNF 1: IFIPLTVQIILKNIIVFALPVHQVRVDIVAMERIHXOETIEHYFEFPAKYEVEISI KOLLSENQPINOEKHOELKKAEELIROAANSQGIYNILLADIAGLVAEVLIVFERGSIITOQYLOSFEALNDITKVEIF	
S76728	ck: 9759 len: 187 hypothetical protein - Synecocystis sp.
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{157}(P)xx(S)(L)xx(V)(T)(R)xx(G)xxx(L)x{114) MLNKSVOILSGVLAALALGTTPAKAEFFORTGNILSLQGHMUYLPDPQMSFDGLAIGR	
PAPPIRINGLM	1:
S76937	ck: 9556 len: 312 hypothetical protein - Synecocystis sp.
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{139}(P)x(S)(L)xx(V)(T)(K)xx(A)xxx(V)x{157) MMKTLITGSGGISGLTFLTLAAYQVTVDFNFEQNLSAECQYETENVINGDKRKEDL 1: VYLIVNDPVYRATDYVVFESHEFNRIHNRADKAYCLMBLEFESMGKPTNV GLEPANLSKLELCALKIKLNFVLENLIGEDPORDIIVSNKILSTGFPMWLSERGIOELIKGTIILRNSVYSNV	
S74351	ck: 4892 len: 333 hypothetical protein sl10071 - Synecchocy
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{100}(A)x(P)(I)xx(I)(S)(R)xx(W)xxx(g)x{21) MRKLIYFLPLBLGKFPVCGGIMALKIKIRIJOECAPTIVYQAREPNMLFLDDLISQGDIGOV 1: VEDLTGIFENKRAYILDSAEYMAOOQVSNGSFLOPEHMACGCCSHVSSNGSLSDY LDPEFCNEIRAGVALDYCDORILGVGRSQRWRVDPMLLAEYRRENTLERLVLDDINEFFDYHHSFQSOTEIPLTORIAWLK	
S74733	ck: 376 len: 156 hypothetical protein sl10931 - Synecchocy
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{95}(I)x(P)(M)xx(L)(T)(K)xx(P)xxx(g)x{45) MASPKRILOKTFIIIASGLAFLGMVTPMLTVLRGNANGREGSQCTPOOPFAADLERLEKAVAG	
S74723	ck: 8425 len: 125 hypothetical protein sl10939 - Synecchocy
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{180}(A)x(T)(L)xx(L)(S)(K)xx(L)xxx(I)x{29) MEILDHYIGNLVSTRFVLSVSULCIILDIYTLQMVWAGDRRTMNTRFRFNSILOFGL	
S75160	ck: 4206 len: 215 hypothetical protein sl11635 - Synecchocy
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{136}(I)x(X)(L)x(V)(I)(R)xx(M)xxx(V)x{43) MDVAFSLTNPVELVIDGEPGLTIAYCARVSSPNOEONSLQLOFCIRBGWASITFEMVDM	

1: MGIDPGLATYRGSPGSGLEPIITTPMDITAPADNITGNHTGWHGVLDIGCGLGDNATILARN
GMDIESLEPATVARELDGTEVEKAFNNVNAQRGGS

C70763 ck: 497 len: 148 i hypothetical protein Rv1558 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{17}(A)x(S)(P)xx(W)(S)(R)xx(A)xxx(M)x{125})
MPLSGEYAPSLDWSREAOADTYMKSGGTEGTOLQGRPVILLITVGAKTKLKRTPLMRVHNDGOYA

C70541 ck: 1907 len: 170 i hypothetical protein Rv1577c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{149}(A)x(S)(L)xx(V)(S)(R)xx(A)xxx(L)x{5})
MAELRSGEGRTVHGTIVPYNEATTVRDFGEFQEMFAPGAFRRSIABRGHKLLVSHDARTRYPV

A70639 ck: 9160 len: 181 i hypothetical protein Rv1957 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{135}(P)x(T)(L)xx(L)(S)(R)xx(P)xxx(G)x{10})
MTDRDADDDLDORVGARLAKAQRIRDLRLTQAAVHRAPAGGLTYDLEFEPAVDADPATISA

RGTP

B70758 ck: 1836 len: 317 i hypothetical protein Rv1996 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{111}(V)x(T)(L)xx(I)(S)(R)xx(L)xxx(L)x{190})
MSAOTNTGIVYGVGGSPCHSTAVEMACQNRNVALRVYVVPVITAPGMAFEYSRFEOAOK
GSPTSGLAIAIAPDEASRRGVLDVLAHAHSDMPDLPFRINAPAIEMRLDEDOEK
MLARRISGMODRFPDVYVHKVYCDRPPARLLELAQTADLVVYVSHRGSGPGLGYSRAVNSGAPYIVARIODPAVPA

A70941 ck: 7719 len: 239 i hypothetical protein Rv2018 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{141}(L)x(T)(L)xx(A)(S)(R)xx(V)xxx(A)x{209})
MAGQDELRDPVLTLLAASRLIVPRATITATMDGIERRRANPANOQGPITLALPHPIGSHN

1: MAGQDELRDPVLTLLAASRLIVPRATITATMDGIERRRANPANOQGPITLALPHPIGSHN
LDNIAIA

D70943 ck: 7125 len: 324 i hypothetical protein Rv2037c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{140}(V)x(T)(A)xx(L)(S)(R)xx(L)xxx(P)x{168})
MALVSTARVLYVGGGSGRIGLVAGVADLAGYFRFPYVAGSSAGAIYASLVAAQLQTAGEPVTRI
SSAIPFEFEPVAVRGATVWDGGLLSNFRPALVDRTEAEWRMPTFCIRLSARGPDP
TRVQGVSVSIGIAIETLVSNODNAYIDPCVTARTIEVPADVSRIDEDITAEOREALYQGFQAGOKFLANMNVADCLADCGPPT

H70863 ck: 4861 len: 419 i hypothetical protein Rv2449c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{83}(V)x(T)(V)xx(V)(T)(R)xx(L)xxx(A)x{320})
MTATPRREDIVYATGCTFVCKTIAEYATAGAGDARTALAGRSQRLVLAAREALGESACWPLTAD
GGSFGCTIASMLEVLSASNDPAPAROLSDPFLSSDRAEDELQPODLPSSRGR
FLABELGAVMTAGTINAPTRIVIRSNALDLAIKRRRIEYSIMSVGSTVLAPVYVGGGVGNAMGGLASRIITLLRGLVAKRVVF

E70572 ck: 8454 len: 273 i hypothetical protein Rv2622 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{185}(P)x(T)(V)xx(W)(S)(H)xx(A)xxx(L)x{12})
MANRGNAGOPLPISDBDDHMOGHMLLRLGRKRVLRPGGVELTRILLARAETDADVLELAPGLD
RALKVNAARPLTVAREMHLLAGLHGLVVEHVTATSMALPOSRVIADEGLGLALRFGAG
NLIIHNAARRRVLIMHRTFRHRERLTAVAIVAKRPVDS

F70572 ck: 9524 len: 297 i hypothetical protein Rv2623 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{110}(V)x(T)(L)xx(M)(S)(R)xx(V)xxx(V)x{181})
MSSGNSGLDITVADIDEPANQAVANNAADBLKRIPLLVLAHNSPEVATMLEPPLPGVILRW
IAFDENSRNRVDVLAHANSVDVSVSMEIRPATOASAEQVLYABRLAGWDERIEN
VALTRVYVNRQAPARQLVORSEEAOLVYVSGSRGIGTACMLVSGEIVAOIARIPVYARESLT

G70965 ck: 4149 len: 177 i hypothetical protein Rv2651c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{156}(A)x(S)(L)xx(V)(S)(R)xx(A)xxx(L)x{5})
MSSITFETALRPGEGRTVYGVIVPYGEVTVIRDIDGDEFREMFAGAFRRSIAERGHKVKLLV

E70885 ck: 9649 len: 308 i hypothetical protein Rv2859c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{56}(L)x(S)(P)xx(A)(S)(R)xx(V)xxx(A)x{236})
MDLSASPSGCDLPASPRLSRPSVSDGDPDLPASPRLSRPSVSDGDPDLPASPRLSPLGA
RGTOVNAVALGCTIHOHLPDILGHSGHAGNGVPTFRIVHTASCRILAEIGESAD
VPCTHHQALDQVGEGLVSAVDVGYIEALDELPGTFVLAVQMPREKSIDRLRFKALVDAASGYAGRSQAEPR

F70669 ck: 5693 len: 418 i hypothetical protein Rv2953 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{83}(L)x(T)(V)xx(Y)(T)(R)xx(L)xxx(A)x{319})
MSPAERFEDVLVGAAGFGSKGLAEHLAHSGSTAIALAGRSSELRGVRMGLGNADWPLI
QRMVSGSVATYSEANRIRASDPDEARLVTIDYTLTTRGAEPELGADPDLRBP
RDLAPELAGFTWGGFVOAPEFNTRIVARSALOEYVGRFRYSETMSLGKSMAPILAAVGTGACTIGLGNKYEDRLPRRVE

B70651 ck: 3649 len: 118 i hypothetical protein Rv3073c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{92}(V)x(T)(P)xx(A)(T)(R)xx(V)xxx(A)x{10})
MYREPRVAVAAVEYDIDPDGQARVLVDRINPHGIRNDQVGVWICWDVAPSKELRMYNHQ

D70645 ck: 2556 len: 332 i hypothetical protein Rv3131 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{33}(P)x(S)(A)xx(F)(S)(R)xx(M)xxx(G)x{281})
MNTHEPDAETVRVLLVAVRABSIHNTQNRNRVCPSTLSLEFRPDQMLRSTDPDGRLLISC
RELTINSGRGYVAVPARNEPSPDSAPLPGRLAGTQSDVLPADGGAAIL
ALQETEDDRLARLRAGEAASIVLLATAMGLACCPTEPLEIAKTRAVRAVEVGAAGYQOMLLRVGMADINADPLEPPTRRRLS

F70951 ck: 8567 len: 355 i hypothetical protein Rv3200c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{33}(A)x(P)(A)xx(I)(S)(R)xx(V)xxx(A)x{306})
MAGSWRLRLNKLKLAOGVGLVGLRLIPORASAPARVISRVAVAVALLTLAGIYVVRD
VTDSDGVLERAAAGLVTHGADATSGVRLAGTDSVRSDDTAVLVLT
ARRIAPKATIVASIRAEQNHILRQSGADTVVSVSETGRLLGATITTPSVVENIEDLIPBAALAAEEVEQAEVGGSPRLRLR

C70561 ck: 48 len: 350 i hypothetical protein Rv3626c - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{82}(A)x(S)(L)xx(M)(T)(H)xx(A)xxx(G)x{252})
MTGASELTLNIVDWEFASVGRLLRARPSTETRRVDELITVAEKEAPPVADVTGLIA
EGORKALDQILVETLLEGHAAVNDVAGPVVAVATIRRRFDRRHNRQPLORLVRLALGFAKLISOYTRGKAEVDHVVNRA

H70850 ck: 8984 len: 254 i hypothetical protein Rv3912 - Mycobacterium

1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{153}(P)x(T)(V)xx(L)(S)(R)xx(V)xxx(L)x{85})
MSAADKQPDKSDADAPPLTVELLDGLADDAARTARISRVASDPQAOQILBALRVVRD
PGGPIIDPSPRTSCLSGLYPASTVILGAPPLIDIDARVAVILVPAIDPDLAAVFA

VAPRCSADTGLASTVYVRA

A70565 ck: 5027 len: 226 1 probable cutinase precursor - Mycobacterium
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{112}(M)(S)(A)xx(A)(T)(R)xx(L)xxx(S)x{98}
 MTRPQPSGRPRAGAAARSLTSLVAAATLTLTPALAPASACCPDAEYVFAAGTGEPPGIGF
 1:
 ALTPQFSKTINLCNNGDPCSDGNRAHLGYFGMTNQARFVASRI
 D70696 ck: 8388 len: 326 1 probable dtdp-glucose 4 - Mycobacterium tub
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{142}(P)(S)(P)X(X)(S)(K)xx(A)xxx(C)x{168}
 MEILVTGAGFQDSHLESANGHWYLDKSSRNANRNQGRSHDRAAFSGSVTDGOTIDRA
 FGVRKAGRGALIPRLVROGINGEGTLIFGAGSKTRDLYVSDIGVAYNLVLRTP
 1:
 TLNGQAINFASGKDIYRDIVEYADKFGARLHDAHNGEVOEPDILASISIGQPVETWGDIDYIMWADQDPQYEDDGS
 B70986 ck: 9024 len: 203 1 probable isomerase - Mycobacterium tubercul
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{1163}(I)x(P)(V)xx(W)(C)(R)xx(L)xxx(T)x{24}
 MTRSYRAPPRIERYVLLNDRGATVADKATVHTGDTLHLAFSSYFDLHDQLLTRRAATRTM
 1:
 LTKLGPAPQAPVADDCRLFKRAHGN
 H70760 ck: 2319 len: 202 1 probable lipoprotein signal peptidase - Myc
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{42}(V)x(T)(L)xx(V)(T)(K)xx(A)xxx(L)x{144}
 MPDEFTGADPLTSTEEAGAGBNAPAPRRLMLLSAAVVVLLDLYTVAVAVOLLPPGQVST
 1:
 FGFDDTVGRHADGDTVGRKADG
 B70763 ck: 4418 len: 397 1 probable membrane protein - Mycobacterium t
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{141}(P)x(T)(P)xx(I)(S)(H)xx(A)xxx(A)x{240}
 MGSIVTGTGKRVGMVREVEFTIDIGDLOGLASVTLDLKLAIOPLVALLPDETIASQIKRE
 1:
 NATTFDIDGATYDITLIGIALSLIILIMTRISVALVIGVVALSLGASFG
 LSLVWQHLLGIQIYVLAVALLLAVSDPNLLISFEKIGAGINTGIRAMAGTGVYTAAGLVFAATMSSEVSDRLVLCG
 A70614 ck: 5098 len: 287 1 probable mma2 protein - Mycobacterium tube
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{155}(L)x(T)(I)xx(L)(T)(R)xx(M)xxx(G)x{106}
 MWNDLPHEDVQAHYDSDDFRLFLDPTOTYSCAHEREDMTLDEAQIAKIDALGKLGLOPGM
 1:
 VDHGLPLTLMARFKFAVETIFRGGPPILEKVEBOSATGPTIRPROSIOPHYA
 RTLDLWAEALQEHKSEALAIQSEEVYERIKITLGCALFRVGIIDVNOITLAK
 S13338 ck: 4079 len: 256 1 hypothetical protein (insertion sequence IS
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{68}(L)x(P)(L)xx(V)(S)(K)xx(A)xxx(I)x{112}
 MAGVITASESWIAPRSGLSFQSKLTVLRREGADAVRGRWRWSPLEDRALLVAAVPNTLMY
 1:
 GGYPGTGLVYPRRERKFAVETIFRGGPPILEKVEBOSATGPTIRPROSIOPHYA
 LKDDGHHMAGIARHNTLALTG
 PC4110 ck: 2351 len: 137 1 transcription regulator homolog - Streptom
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{57}(A)x(S)(A)xx(F)(S)(R)xx(L)xxx(G)x{64}
 NHAARRLVAAPYRSGGOAOVYPRSVPEVGERFAATREMAHLRUGEPITLLELARAASARFST
 1:
 PN0644 ck: 9954 len: 66 1 hypothetical protein 66 - Streptomycetes coel
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C

1:
 x{127}(P)x(P)(V)xx(A)(T)(R)xx(V)xxx(I)x{231}
 MAKVTRDDVARIAGTSTAVSYVINNGPRVAPATREVRVLAHKLGTFRPDRVQAAMASRR
 S44233 ck: 6765 len: 358 1 strp protein - Streptomycetes glaucescens
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{136}(P)x(T)(P)xx(V)(S)(K)xx(V)xxx(L)x{206}
 MLIITLGHQYLGIVTMAPVLTAAHGOVTGLDGSFVRLGALADVDPVGLALDLRDVYDTL
 1:
 RSDIYLNINVGRAVAGSRYTLLSGDTGRTVLAARAVLAGLAAPATVYARA
 ENIGTEENKRTVAIEAALAAVAGSVLEISGTDNDPSRYDFTARTELGRATWTPDGAOLAREYRAGLTAAFEHDF
 S17674 ck: 4817 len: 291 1 serine-type D-Ala-D-Ala carboxypeptidase
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{23}(A)x(P)(A)xx(V)(T)(K)xx(I)xxx(G)x{252}
 MRIRRAATVITLTKALTAAGTLGATPATVATKRTIAAVGAYAMNNGTGLTYTKAADTRRSTG
 1:
 ANYSTERHLTKIASSKMNSTFRIVYVTKATKATVTKGTSITNTKNTNGLLS
 SYSGAIGVTKGSGPEAKCYLFAAATRGKTYIGTIVLASTIPARESDARKINNGAL
 S17717 ck: 5969 len: 280 1 rRNA methylase - Streptomycetes tenebraritu
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{40}(A)x(S)(A)xx(A)(S)(R)xx(L)xxx(V)x{224}
 MPHPARGDPDPDPRHAEVDAVRSRRSKNKKSAIPITKIVKFIPIFLITVGLMGCTQ
 1:
 VANDVRVRLMTGVGEVETDVTLLTLKTVPCIEAOGRGQMDLIDALISPLVAVSF
 TKSIGQSKGMFNTYSANFDAMLENRPDVEQLERNELYFVAKNA
 C64251 ck: 2836 len: 385 1 hypothetical protein homolog MG464 - Myc
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{128}(V)x(P)(I)xx(A)(T)(R)xx(L)xxx(V)x{241}
 MPKLAQTNKEIKTTPNPWSAAVVENNNKNEKKSALPITKIVKFIPIFLITVGLMGCTQ
 1:
 KMNVEQKIAENAKKYGALDLOSKRNROLEINIKSNHNKSSAAVQVFTLPI
 FLIIRYVITLPIKAILLENFMDLSKVLTEIFENFTTGMPIFLIYVLPVOLFQKLPQWMAKRNENAKH5QKSIQOLN
 I64230 ck: 1048 len: 248 1 sensory rhodopsin II transducer homolog
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{46}(A)x(S)(A)xx(F)(S)(K)xx(L)xxx(L)x{186}
 MLVFEVILFGITLALVRYSTAAALVYQANRISIDSSFNKAKAFANALANSANQESKSTNNLD
 1:
 DVTMTTYAVSGGHALILITTYVFSFLISKKGGLIFRSFISTEQDLADHVNDIIR
 YPDLSEEVITLADHD
 S73488 ck: 5866 len: 385 1 hypothetical protein K05_ort385 - Mycopl
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{128}(V)x(P)(I)xx(A)(T)(R)xx(L)xxx(P)x{241}
 MPLNKRNKHEKLTINPNPSAAVVENNNKNEKKSALPITKIVKFIPIFLITVGLMGCTQ
 1:
 KMNVEQKIAENAKKYGALDLOSKRNROLEINIKSNHNKSSAAVQVFTLPI
 FLIIRYVITLPIKAILLENFMDLSKVLTEIFENFTTGMPIFLIYVLPVOLFQKLPQWMAKRNENAKH5QKSIQOLN
 A36149 ck: 5138 len: 237 1 spiralin - Spiroplasma melliferum (SGC3)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{40}(A)x(T)(P)xx(V)(T)(K)xx(L)xxx(L)x{181}
 SILAVGSAVFTSVACNCKEINSSKNSRSLTAAPRTAASAPKAVTPEIKTALANVKA
 1:
 IONIAFNAGASDFEITNNGAEDYEAAREVEYVAKANDSANISGOFRKAKVATL
 APTF
 H69019 ck: 3145 len: 157 1 conserved hypothetical protein MTH1148 -
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
 x{3}(V)x(T)(A)xx(F)(T)(H)xx(V)xxx(I)x{136}
 MIRVATAECFTIGFARHIAHYSMGYPCGSWSVSDVVLVAGLFIPTLGSIRSLIKFEPPEP

D69049	ck: 1234	len: 355	i conserved hypothetical protein MTH1371 - Me
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{153})(V)x(T)(I)xx(A)(S)(R)xx(A)xxx(I)x{186) NMLNVAAEMELGSLTRSRKRLFIPLFOIALVLTITLFMEGSFIDVLESEALTPSLTGFPAELGVDF 1: NMVYDSIVGERKERETGELLAMPVRREIIILGSCISVTIVLADQGMWMLLLSG FMHNPILAFTVVAASAPVAGLALISVARNYREAIGITLVLAIIISAALVLAPALAMVGSSGVSPTMTLIKMSIQAMGLPSTF		
D69065	ck: 9991	len: 249	i conserved hypothetical protein MTH1489 - Me
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{86})(V)x(T)(I)xx(A)(S)(R)xx(A)xxx(I)x{147) 1: MRLGSTLAFMEBLERLILNRKADDFEVELLECCEGPYPRRLTDODGSLEVFESFDLEVLHAAPT ANTTGRLDDELFEIKRTAHAYHISDNNGKRDOHPJLGEVTVDLKLGSIERGYEINLS YDGVIRSRRIILEEVRR		
G64362	ck: 6767	len: 406	i 2-isopropylmalate synthase (EC 4.1.3.12) -
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{160})(A)x(S)(A)xx(A)(T)(R)xx(I)xxx(I)x{230) 1: MKTAVYMNDLFFENSKRACPYNPKLDKDIYITTLDRDECGTPGCFCREQKLIARKELBELG VHKAAEBEGADRVIHIAOTTCGAPOSHEVCCKLRKAHGHCANDSGPAVI NSIYGLIGAKAAVSTVNGIGERAGNAALELMALTVALVDVLDLNEVLPELCRMVEYSGIKMKPKPIYGLVAFHESGHVDA		
C64456	ck: 7004	len: 167	i hypothetical protein homolog MJ1251 - Meth
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{3})(V)x(T)(A)xx(T)(T)(H)xx(I)xxx(I)x{148) 1: MITVAACEFHNTANGITIHKAAGATDEDEFKLTSEBDUKMAKNVRYISMFPSTIGYEKLDI		
D64423	ck: 9981	len: 329	i hypothetical protein MJ0988 - Methanococcus
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{47})(A)x(S)(V)xx(L)(S)(R)xx(L)xxx(G)x{266) 1: MBLUBRDVLELCHLNADSPVACSVALKYLALQSOLPNKFRISADSYSKLSRNILMEIGERY KOISFORKILLLOSSESVKRTHAKCSMMELTFREDKRLIALSHVSSHEASCART IVSIGADVAFVAAVERKEKIRVASCRKHVSXYVHLGNLMERIKGELGGSGHSEAGLINAPYDKSKSERVIREVINLICRKRFVE		
D64452	ck: 2920	len: 299	i hypothetical protein MJ1221 - Methanococcus
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{108})(A)x(T)(M)xx(I)(T)(K)xx(I)xxx(I)x{175) 1: MMFGSRMYKMIIIPKRLNTVEBILKNNANVSIIEBLKTSIEDGIIITCNDAADAERIVLEI LLCGSIIVCAEFISFLFVSKRELIVLOLYSTSLISAIVAIIAGISGMSIASAGE YEIIGVTIVDSILPALMGMALATMDLYITYTFLLAINIVLDDVGYGIGLKRYKGINOKIRKY		
F69396	ck: 9110	len: 437	i acyl-CoA dehydrogenase, short chain-specific
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{189})(V)x(S)(A)xx(A)(T)(H)xx(I)xxx(I)x{233) 1: MYMPKRYDDELDESEEDRPKEVHFAPFIRPASIELDRAPPERVKPSPCFYKWQIKRKGX DEMVISGOSAMWSAVATNCGHMOICGSRBIAGCEIYIPADGVKKGRPAD MLGMHDPQGBELFFDGVRIFREHHVVVAPGFYGVFFDQLCLITSCMGAFVAGLARACFEELHYANQRQGVPIYVKHNKVLXE		
G69440	ck: 1599	len: 136	i conserved hypothetical protein AF1528 - Arc
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{7})(L)x(T)(L)xx(X)(S)(K)xx(P)XXX(V)x{113) 1: MDVNIALQTLLTKMSKPVIMLVSLITSNAPSLSLPROFSRGVEFINNLIRSKLSGFAINSFLIHF		
A69470	ck: 6749	len: 61	i hypothetical protein AFI1762 - Archaeoglobus
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24})(M)x(S)(M)xx(L)(T)(R)xx(A)xxx(V)x{21) 1:		
D69509	ck: 4730	len: 270	i hypothetical protein Arz077 - Archaeoglo
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{135})(V)x(P)(A)xx(M)(S)(R)xx(L)xxx(T)x{119) 1: MKKLLIATLAEVLISGSSGKREYRGARGAVANADNSSYIGFECBPENTLYCTANGDSTGVIV LDITPAEFVLISAADSGSFATHHPPAGSGCTKIWDVYVEGEFMDVATKRN PAGROETSSTGSYNLDGAELIRGLSVNPIIVKAVR		
T08352	ck: 7002	len: 148	i hypothetical protein H1537 - Halobacteri
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{40})(L)x(P)(M)xx(I)(T)(H)xx(L)xxx(L)x{92) 1: MSNSPEITHDFDAMEFPGVEDGCCRIRAIADRBERSYTGELAPLRITHOGLTLDAGCWIG		
S73242	ck: 4681	len: 291	i hypothetical protein 38 - red alga (Porp
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{116})(L)x(T)(A)xx(A)(S)(R)xx(I)xxx(I)x{159) 1: MTFASFKEIRIKLKFEDPKYKSELIOEBELAVORLUOVNRBRPATLMAGTIQPLWLVEFY VTMLSLSTLFTLGHLELLAILVYNLAFPSSTLAPVPMPLDILSNLSDPS YAIGIRYISNMDMFSTESIWSGDISISQITISLLFDVIGAVIYNNILKARN		
S20471	ck: 4134	len: 86	i class V zygote-specific protein - Chlamy
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{131})(L)x(T)(A)xx(A)(T)(H)xx(P)xxx(G)x{157) 1: MKGRFKSLALIALDATFSAHSATHGPIATIGICTGCNALVACIYAAGTEFVPSMGAIIRA		
S25969	ck: 7679	len: 69	i hypothetical protein 69 - liverwort (Mar
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{20})(L)x(P)(A)xx(T)(T)(H)xx(P)xxx(G)x{133) 1: NASLETRIOTFLYPAHTFIYPAHTFHSHPGAFPOALHLPGLGLEFERNAHCHIFDNN		
T01531	ck: 1750	len: 289	i hypothetical protein A_Ig005110.4 - Arab
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{131})(P)x(P)(A)xx(L)(T)(R)xx(A)xxx(A)x{142) 1: MVHAKSATIPIPERKRIKRVHVDORLQAFORTFNALITYLMLPFGLIISIRVYNLPLPSR LFAELSDRIVPVAMNOKOGMFGNFATIRVGVKVEYEPFEMNRPSEYEAFTPEADPEE MTVNGGKRTPIEVANVOYKIVAGVIGFECTELTRKDXYLLGNDGKVESINWTKR		
T05405	ck: 4069	len: 222	i hypothetical protein F10M6.130 - Arabido
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{170})(L)x(P)(L)xx(L)(S)(H)xx(V)xxx(G)x{36) 1: MTGIVASYLFSVARNNEELILRWVEODEDHWCOCBEETLICYAMDOLDLGRGSHSYTAGRC SHELVRFGSTIEIALGFGSGNKKLPFKEDLMKNSNBIDLGS		
T00559	ck: 9966	len: 296	i hypothetical protein F12L6.15 - Arabidopp
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{144})(L)x(S)(V)xx(V)(T)(H)xx(V)xxx(V)x{136) 1: MEHRKPDIISCDIBELLIXTISLIFPSRARPLVSTRFSVMNOALLVAHSHNGSIDISHRA TSLTIDSPRTLHLHSISGCPKRLIETIISFKTHTHGOISLPKRIEHNLTRKAV FDVTOGSTYNNALDIGPLSLITKNSQSLIDCRMWFETKLVSFFCSKILIRLYIVTNKR		
T01617	ck: 1414	len: 392	i hypothetical protein F19F24.9 - Arabidopp
1	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{116})(L)x(T)(I)xx(V)(S)(H)xx(V)xxx(L)x{260) 1: MKTRISNRKRLFESFRBSDBSENSSPLAASSTCVCEFOCTDOPRKPKCESFOEIYD		

1 709662 ck: 3889 len: 101 i multicatalytic endopeptidase complex chain
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{177}(V)X(S)(L)XX(X)(L)(T)(K)XX(V)XXX(V)X(8)
 1: CDPSONYGMKAGAGIAGNANOAGSLKDDYKDDITREAVNALAKVSKMDRSLSDLELA
 1 523774 ck: 5889 len: 402 i triose phosphate/3-phosphoglycerate/phosphat
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{83}(V)X(P)(V)X(X)(S)(R)XX(A)XXX(G)X{301}
 1: MESRVSRATLISLPLTLKRLPLANSLPSKSGSDGNLWGRQLPELCSVLKKGAS
 TSNSFAVAASFTIRKALPEFFENAAQSOLGOSITPLTLALPAAVIGVSMAS
 LTELSPNMLGFTISAMISNISFTYRSIKKATDMDSNTIYAVISIALVICPALLIEGPTLKTGFNDALAKVGLKVFSDLEFW
 1 522500 ck: 9347 len: 256 i embryonic protein ecp31 - carrot
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{185}(A)X(T)(M)XX(A)(S)(K)XX(L)XXX(L)X{155}
 1: MSQOQRRRPOEOPIKYGDFVDSGOLSSOPVAPNDASAMQAAENNVLTGTOKGCPASVMOGASAAA
 AOSADYNARTMNVASKTRIGDVLADRSIRLAEDEKAVREDAGVYGAVERNPNEM
 MTPGCVASMSMAAARLNODPTF
 1 716984 ck: 4925 len: 165 i transcription factor homolog BTF3 - curled-
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{108}(L)X(S)(L)XX(A)(S)(R)XX(P)XXX(A)X{41}
 1: MNVEXLRKMGASVRTGKGKGMKRAKAAVKTITTDKRLKSLKRGVAIRAEVNIFFKEDVYI
 1 547086 ck: 1305 len: 263 i p17a protein - rice
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{52}(V)X(S)(L)XX(Y)(S)(R)XX(L)XXX(A)X{195}
 1: MEDGGHFEVHOLGAGGVKRVYALRAAGRAVALDMAAGAHAPADEVGSLEESRPLDA
 RRGNYIDDPIMKDDTLTLEGNYGVSQKRVFVAMDDASSDEDMRWITLDSPGVEY
 BELAGADHMAKCSKPRELCDDLRLTAAYD
 1 514884 ck: 7397 len: 146 i hypothetical protein 7 - yeast (Hansenula B
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{44}(L)X(S)(L)XX(Y)(T)(R)XX(A)XXX(T)X{86}
 1: MGSTSGCLYSIFISALAKSLKDLRELQOGGEVPELTALASLKDALDSLRYTRARAGDRLIDALA
 1 543448 ck: 5386 len: 215 i FUN54 protein precursor - yeast (Saccharomy
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{33}(A)X(S)(L)XX(E)(S)(K)XX(L)XXX(M)X{167}
 1: MKRSTALSPFVYLLALVNSVAASSYAPVAISLPASKECLYIDMTYEDDSLAVGQVLTGSN
 ESRLLWLSLIIIIINAVISIAQVLLIOLFTGRKNTY
 1 519425 ck: 1969 len: 317 i hypothetical protein YCR015c - yeast (Sacch
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{93}(L)X(S)(V)X(X)(T)(K)XX(1)XXX(1)X{208}
 1: MKTIIISDEDEITRVDICTIAKLPYLINPRKPEMGHFTKYMDGYKRYNGTRSLPLSSGY
 LKRVSDKCSQSYNGEFDRLTLTGSDKRVILIEIDKIDSGCNKSGNSYVIGDS
 ETDLSILHPSITNGVLLINPOENPSKFIETIKIIGIPKDKISSFEADNGPAMVLFCEKEGKGAVALVKSVDLSKDLIMQVTKM
 1 570131 ck: 1796 len: 365 i hypothetical protein YDR273w - yeast (Sacch
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{137}(P)X(S)(V)X(X)(S)(L)XX(V)XXX(L)X{212}
 1: MGRKNKGEENNAKSTFLKVENIKNTNGSLPDSODYTNVEKESSPRTDPLTLKHEVNTKIDS
 KIESELISLAHIGNISITPGSNRNEVAETBRVAEAGCNTILISNDSTLKEKRGK
 EEEKEKGEKGVNSLKGAAVYVAKSLKNNRIPTVAKNEPSNNLFDVLNCDSEDEEDQVEINTSNOEKKNNGCTEYVPEAORSA
 1 569699 ck: 3355 len: 374 i hypothetical protein YDR415c - yeast (Sa
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{127}(A)X(T)(1)XX(I)(T)(K)XX(P)XXX(L)X{231}
 1: MRIOSTFLFNLVNAIXYFEPFLVLDVGENEWEVPESEKLNLRGKVKFPDYTHHTSELP
 ODSINILLPSMAAPGADDGSGVITNMEALRGLTEFLKRGFRNNTEFEHTSA
 EESGELSDLVFTIAKQKNAHVRALQDDMTGVSDEDEHGVITDYTPALDITKILINSYLSIPYDQTQCYACSDHGSAT
 1 550587 ck: 5798 len: 128 i hypothetical protein YER084w - yeast (Sa
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{49}(L)X(P)(P)XX(P)(T)(K)XX(M)XXX(S)X{63}
 1: MBLICVPHYSNLFMEFLFECPSKRRARHPRKFTLLCYKSNHLLPKLLPPLFTKRWMLN
 1 564117 ck: 6897 len: 107 i hypothetical protein YGH109w - yeast (Sa
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{34}(I)X(S)(L)XX(Y)(C)(R)XX(L)XXX(S)X{157}
 1: MAONPLADIOVYKRYKAKRRMEGOKKNCNCTIAYIDSLOYCRSLSHKSCFPFSQAFAS
 1 546801 ck: 5569 len: 429 i hypothetical protein YK027w homolog - Y
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{196}(I)X(S)(M)XX(A)(T)(K)XX(P)XXX(S)X{217}
 1: MANTWRLIATALISVPSOLASIAKREKXLSAANKNTVSRFQYDDHLFREOLARNYAF
 NIESKVDLFEYAHNRKIDVYSSMGVATKSDPTVSINDISMTFEDPISHCYRKLK
 KRGIATISVYFNSHMLDPARDLISLIDCEHNAIVADRLHLPBLGTMGILGSLATWILTKVSGYPMKENEVANKRLKY
 1 564829 ck: 7470 len: 336 i hypothetical protein YER007w - yeast (Sa
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{124}(I)X(S)(A)X(I)(C)(H)XX(A)XXX(A)X{296}
 1: MEVHEOVASVNTDAAKAYLLOLILSARGCENALILALMRLETASTLNTWMSIQGVWDK
 LETSIIVKEVNRILVATGDSNLAWKRTKFTVSGTINLFOELATATIDELLR
 LCEILKWFYRQEGKFIIDLCIALEBYLSTMYNLTCONCHKLAIOGVRCNESCNEENETGENSLSQIWHVOCFKHYTHVS
 1 568481 ck: 4630 len: 254 i hypothetical protein YLR165c - yeast (Sa
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{199}(L)X(S)(A)XX(I)(T)(K)XX(L)XXX(T)X{39}
 1: MSLKQIPIIENHTYIVNKPGLPISQPPDCRTWORTHNDLPTPLERFKAIYSHREVEL
 VNEGCEITLPIKREVLLENMDQ
 1 569848 ck: 7663 len: 105 i hypothetical protein YML101c-a - yeast (S
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{35}(M)X(T)(I)XX(A)(S)(R)XX(L)XXX(A)X{54}
 1: MYRIKIKNHNKSTGVCMPLFLSLLOGIFFTGSMFTIPASRWFALATALLSSVSSGSAC
 1 577699 ck: 7370 len: 238 i inner cell wall mannoprotein ICWP - yeas
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{9}(V)X(S)(L)XX(L)(S)(K)XX(L)XXX(P)X{121}
 1: MRATPLTSSVSLILSKLEVLAHPALPCLVACVGOVSSSTCDLSLNOVOTCYCHEMSAVYKCL
 QASSTSEAHSSAASSTVSGEFTVSALPSTAVISIFSESGSNVLEAGSVETIAV
 AAMLI
 1 540550 ck: 166 len: 441 i peroxisomal membrane protein PAS3 - yeas
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{144}(V)X(S)(L)XX(L)(T)(R)XX(L)XXX(T)X{280}
 1: MAPQSRSLORHRCGLVLSUGIDALFLTGSLVYVFAKRWLTKOOLRITBOHFTKEQIKRR
 NCKNLORRYNNVYTSWSPDEKADAMVMAKRSKKEQEVYINDEQALSLSMWI

LINKGLSYNEIITNOILEEDGHPDRLTLEEFSSRLNIPFNTNSQIFQONNNNTSILLPKDSSGQEFLLSQTLADALTSFHSN

1
S67699 ck: 4697 len: 193 i probable membrane protein ydl151c - yeast
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{16}(L)X(S)(P)X(T)(T)(K)X(A)XXX(P)X(101)
MSTSLFSLSPSSSSMRLNRSNFPILNFDLDTPLSSSSASFSLSLPSLIVSRPCTRDI
RPFLLILLNECPA

1
S60916 ck: 7226 len: 139 i probable membrane protein YN266w - yeast
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{82}(A)X(S)(L)XX(L)(S)(K)X(I)XXX(S)X(41)
MMLNHTYLLSYFLRKASNRFNFSSSSSFSCSEFLVFLFVFFSDCFISITSLISFGLISFLIF
MTEALEIVLLVIOSTLOXICRQICIAFLIPFLGLYAFDELFLYRMILYSOMENYKRLGRSKTN

1
S66918 ck: 8726 len: 157 i probable membrane protein YOR044w - yeast
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{112}(I)X(S)(L)XX(I)(C)(R)X(I)XXX(L)X(129)
MTEALEIVLLVIOSTLOXICRQICIAFLIPFLGLYAFDELFLYRMILYSOMENYKRLGRSKTN

1
JT0589 ck: 609 len: 331 i pectinesterase (EC 3.1.1.11) precursor - As
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{11}(A)X(T)(A)X(S)(S)(R)X(A)XXX(R)X(304)
MVKSTLASVFAFATLAKSRMRABSGALVYAKSGDDITISAVALDSTISTETOTIFIEESYDE
EGAVDEIFSCGHRARMEHEDIRLVGSPASITVANGRSSESDSTYIKHSITAKM
DGNVSSGTYLGRPMPSOTARVCFOKTSMTDVINHLGTEMSISTPNTENVTEIGNTGCAEGRANFSELEPTITWLLGSM

1
S03833 ck: 3269 len: 319 i hypothetical protein 1 - chestnut blight fu
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{62}(L)X(T)(P)X(V)(C)(K)X(V)XXX(L)X(241)
MAOLRRPSOLVISESVPTVDPFVSRTEEVVAGCITLMEWRDSCGDPVPSLHSGDLRLRTH
TTGCMCVADVLLTLOWVGRSFGFSGFQLEKSAVDHVVADAYOGEODGALFY
AIIGLAEKDPILARIGRLNPLAEPASALREVEPQVTRRKSTRTMGTRDPTIVFMOGWYGHQHPGSCSGYVEERQFRPS

1
A34051 ck: 7750 len: 201 i NADH dehydrogenase (ubiquinone) (EC 1.6.5.3
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{1106}(L)X(S)(V)XX(A)(T)(H)X(V)XXX(G)X(79)
MASRNVGYGVKITAGGVVSGKITVQSGWERRIRAFALDPRNSGVPLVPRNRPSPGSLDPL
MDVYKIQIAEPESTSEATFCRSFS

1
S16556 ck: 4238 len: 289 i hypothetical protein 1 - Podospora anserina
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{1172}(A)X(P)(I)X(V)(C)(H)X(A)XXX(I)X(101)
MSEFEGVAAALVAGLFNNCGVDFEYVOLFGRFGRDYERCOQLRLDAKARLSWGAAYVINDPF
AVCHKLAIEIEVEEEDASLTLLIKDAGSIDAMDAADAKIDAIIVGRNSAKDIRT
EERARVOLGNVVTAAALHGRISDQITNSVETVYKGSRYLLIGNEXGKGFWND

1
S16557 ck: 4187 len: 289 i hypothetical protein 2 - Podospora anserina
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{1722}(A)X(P)(I)X(V)(C)(H)X(A)XXX(I)X(101)
MSEFEGVAAALVAGLFNNCGVDFEYVOLFGRFGRDYERCOQLRLDAKARLSWGAAYVINDPF
AVCHKLAIEIEVEEEDASLTLLIKDAGSIDAMDAADAKIDAIIVGRNSAKDIRT
EERARVOLGNVVTAAALHGRISDQITNSVETVYKGSRYLLIGNEXGKGFWND

1
JN0451 ck: 2342 len: 312 i phosphoribosylanthranilate isomerase (EC 5.
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{62}(L)X(P)(L)X(X)(S)(K)X(X)XXX(I)X(234)

1
GIDLVLHGDPEPMANFIPVAVKVERVSPETVIRGRGNQALILADAGA
SOGGGEGRAPFMEHAKRLIOSGEVSGHPLVILAGLIPENVALNRLVAFGVMSANGSKREGRSRLRLRS

1
B44418 ck: 3830 len: 412 i surface antigen - Trypanosoma brucei
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{112}(I)X(T)(V)X(P)(T)(R)X(P)XXX(T)X(284)
MCIOLVSDDEEFKAVACLOCLILGIGVFLIGVFLSDPDRDNKRKAVSAFDPLEBSW
EKGVCNKEVYLSKLYVAASDGGKGEYREESTOYASLSEFGLHLSOGYEAVPDQXVO
VOLYBEGPFLALBEIMGESEFVPRMTGICVAGSILLILEIACVAVCECLKRGSSNDTGDPTDQGGSPYTGOS

1
T14622 ck: 9305 len: 425 i hypothetical protein - Trypanosoma cruzi
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{143}(V)X(T)(L)XX(L)(T)(R)X(M)XXX(A)X(266)
MRLNPALEILLNOCGLNGRSPRSPSSGGRKRGSGRISTDARLKRKAVIADALSTIVG
VHRLCAAPLYVPIVIGNITITTSKSGKATLBNQVLRNDPCHASMGERRSGAT
QYTFIMVRFDHTHRAVSLDKFVRSVSRMPLNSITIAEVEVASFLIAAIIITRSCDYFFIKAVRRRLSALNNGIYQETSP

1
T15458 ck: 2648 len: 331 i hypothetical protein C08A9.7 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{102}(L)X(S)(V)X(T)(T)(K)X(I)XXX(L)X(213)
MVSATRVPRRSSTTSATVAAOQRTSPMLPASFPITMDEVLEKNEFEVNVASKDIAMKALIT
FDGLFEVENNEVHNLPMMEKRVESANLHDEQANNEFDOTENRNMGAATDSE
SQIRHETSSPMNAQISINASIEYENVPQIDTADNIGOVKOLFVDHDBRANFREVLTFTVLELDRPFTNAGVFEDMSLS

1
T15482 ck: 121 len: 217 i hypothetical protein C10A4.3 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{96}(L)X(S)(L)XX(P)(C)(K)X(L)XXX(M)X(105)
MNTSRSPILLMOCSFISICITLTAANTESINHRILIPROCTCNPQHEPPYOILNHLISOI
IEVRYDQPLPWTEDLDLPAVSYLFFPDSQLLKREDC

1
T15182 ck: 9118 len: 150 i hypothetical protein C18E3.4 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{78}(I)X(T)(L)XX(L)(S)(R)X(L)XXX(L)X(56)
MSLTMDNCKSRMKYITVATTITLDELITVIGACVLCSTVSVSICNOQGFVYIDKILLYI

1
T15715 ck: 4709 len: 280 i hypothetical protein C30G12.2 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{197}(M)X(T)(A)X(M)(S)(R)X(L)XXX(A)X(67)
MADVADPEVESRORPKNIMITGANGRIGLVAKHLEVDGILILATCRNEKADDELIALK
GCDFGPKRFEHYFLVNSOEMTTAYRMSKALNFAROSFEMLSKYHILVTSFCPG
MWRIDMGDADLDVNESTKRLTSLANILRLNNGLYDFRLHIPN

1
T15729 ck: 770 len: 304 i hypothetical protein C31H1.6 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{155}(A)X(T)(M)X(V)(T)(K)X(L)XXX(I)X(233)
MSDESGISDPSKSPASPEDEKCKSNRNEPSTYKGVNDKADKSLDILEBVVNMKATMEDVT
NHIGTSEELTWGVDSCVYKPFVPAEYVTEIDEDPILRISGKLYVT
ITNLEKNSLVITTEGNTCDILRGMEDYAKFRIDGPAIYTERGCIIRFGLEKVRINRESLNSSRX

1
T16486 ck: 1289 len: 374 i hypothetical protein F56D3.1 - Caenorhab
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{144}(L)X(P)(A)X(X)(I)X(X)X(215)
MKRLITAFMPLGVNKAQOEKVEKALSEADONENKRIITVIGITDGRRLREVDGQIEV
APPRVIAPOOPDFRIFQICOPGDFGPPGAPFOAFTAPFOAFT
RPPRPDPNATRBACOPAPFTTPARFEPGOSNLRSQCOASIFITSPISGPSRLFTFRAVAVYDQCARICHEFNACSN

[illegible]

1	G71287	ck: 2410	len: 365	1	conserved hypothetical protein TP0730 - syf
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(175)(L)x(T)(L)xx(A)(T)(H)xx(L)xxx(S)x(114) MARASSSELSMHSBPSLPSAHCAALRRKTCQSGSDASFQKAEKSTLPKPARPFPPVYFH TLNTATHTILIRASAPIVGFLKFKHALGSLYFVAGLSYFLTFDFIDGKIKIARAR RETSSVGETLADSDYALIGLISALYOSGVPLMFVILITRLSLQIVACVYALFGHPMTGSTAGKATVATMLTYLLEIARLLI
1	J00857	ck: 6140	len: 169	1	hypothetical 18.1k protein - Escherichia cc
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1124)(L)x(S)(L)xx(V)(T)(R)xx(A)xxx(V)x(129) MFDYQVSKHPHFDEACRAFLRNHLVQLAERAGMNVOLIRKLNPSOPHLLTAPEIMLTDLIEDS
1	T10288	ck: 562	len: 298	1	hypothetical protein 19 - Orygia pseudocysg
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(46)(A)x(S)(L)xx(L)(S)(H)xx(A)xxx(G)x(236) MLVOVNFIDLVLAHLAFGLGCSFAFYFALMAYVIAHAFLELEDSANSHLSHLAFLGPIYI VRHDEPTVRNONGALITMALLAIYVHCHNMKVQKRETRRKRANTITLMEI EKEDYAEHEEHENRSMWLSARRARNPHNPFECROTOTQVNSRLTFQHRARAROVPTAE
1	G72757	ck: 9964	len: 175	1	hypothetical protein APE0052 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(37)(L)x(S)(V)xx(A)(S)(R)xx(V)xxx(V)x(122) MPSALRTITIAAVNLVAFIAGVAYIYSEYSDVSTQSVTRASRVTVQAGVAYLGYGTATVYG
1	C72759	ck: 3453	len: 114	1	hypothetical protein APE0064 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(A)x(P)(L)xx(V)(T)(R)xx(A)xxx(A)x(189) MTSLGACSIASPIILWVRSTNAPEAKRPAASSIRAASSGVCVSSSTALEPRARALITSS
1	A72773	ck: 7998	len: 129	1	hypothetical protein APE0172 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(82)(L)x(S)(A)xx(L)(T)(R)xx(L)xxx(T)x(31) MSPULSIINGADARVSSISLSTGFANFRVTCOMNVGAIAYMAMVSSGSGTIVLTSTAMWLG
1	C72775	ck: 1302	len: 100	1	hypothetical protein APE0189 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(24)(P)x(S)(L)xx(V)(S)(R)xx(P)xxx(S)x(60) MLFIRSRISDIIFDLKLAIVLIIIPMSISPVSSRPIITCSTLIVIMMAASSVASCFRDILA
1	C72784	ck: 9325	len: 313	1	hypothetical protein APE0258 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1152)(V)x(P)(A)xx(M)(T)(R)xx(A)xxx(I)x(145) MMELVGSLLDVIVEVSRALGKDFOGGLKRSGLNORAKVAELTSHPEILBAKRVTEVRESYKH ERDVEGVAAVREFLEKEFEKARVGTIGTGNNSAGTIGVIVENGRRLVSLSP VHAALVPIDKIVBSVMDAVNALVQAAGFMMPYISILITSPATGIDIEQIKVLGAHGPREVHVLLDNGRMAKASHPF
1	H72721	ck: 4962	len: 168	1	hypothetical protein APE0313 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(8)(L)x(P)(L)xx(V)(S)(H)xx(P)xxx(V)x(144) MMSGVELVLRPIGVSHGLPDSTVRENIGVGEIVLEPIETAPGLDGLAGFSHITVAYLHKSXGR
1	H72731	ck: 2984	len: 172	1	hypothetical protein APE0393 - Aeropyrum pe
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(65)(L)x(P)(A)xx(A)(S)(H)xx(I)xxx(G)x(91) MGSTLSSSIPSSSVLAISSAAALAIPSVWARDLTSRAPLTKPENTRALTLTWGRSLLEVA
1	G72736	ck: 2336	len: 180	1	hypothetical protein APE0429 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(26)(L)x(S)(V)xx(V)(T)(R)xx(V)xxx(P)x(138) MSASVARRRRESKKEGEGEKSKDLKIISEVETVRDKVVKNPROMALLYLIDRLQPIHERTLQ
1	D72737	ck: 7680	len: 120	1	hypothetical protein APE0434 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(159)(P)x(P)(P)xx(L)(S)(R)xx(I)xxx(V)x(45) MRSPFWSSYLVFNLISGENLGGCGTATGPQVQASLTSVILPHLHFRSHCNLASPO
1	A72741	ck: 338	len: 160	1	hypothetical protein APE0458 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(9)(A)x(T)(P)xx(L)(T)(R)xx(L)xxx(L)x(135) MVGGAULTAETPAPITRPLDLKDLSTQARASLWGLSVLSTSTPILSPPLTWPTLISS
1	G72640	ck: 9276	len: 147	1	hypothetical protein APE0559 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(62)(L)x(P)(P)xx(F)(T)(H)xx(A)xxx(P)x(59) METTPYRAASTASTRTPPGSHGASMTPTPLCLAMYSISLTLALAGSLSSWSLGTREPEPL
1	A72651	ck: 639	len: 119	1	hypothetical protein APE0638 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(102)(L)x(P)(L)xx(A)(S)(R)xx(A)xxx(T)x(40) MMTRLDPPAGAGGIYISRCLSATPFTSLAGFRASARTSKALMGFWPPLYTSTVANYXS
1	D72672	ck: 4161	len: 129	1	hypothetical protein APE0802 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(1102)(L)x(P)(L)xx(A)(S)(R)xx(A)xxx(T)x(111) MPLIKSLPSOLSCSLNLSDEMSNFTSRSPSSIPSPENGLHGLVTLNPLGLGLMLITIN
1	E72679	ck: 8681	len: 234	1	hypothetical protein APE0857 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(173)(V)x(S)(L)xx(A)(S)(R)xx(P)xxx(G)x(45) MLLRNPKGQVETGPGPGGODEGVVHPPIVRAPOBLAKGAEMDLVEVPDAGAEIWEWH GPASRGCPULSGAIVDRPORAAGRVYTGHRPPGSRVALLPHYNGPVAIVYQRG
1	F72685	ck: 9005	len: 228	1	hypothetical protein APE0902 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(170)(A)x(S)(V)xx(L)(T)(R)xx(L)xxx(V)x(142) MEALYRSVEOLVNLNMDLDTLGGSSDRIQVLEPIYVLENAKSLSRPTAGIARKIIGQKHLEL IVEEVLRSIIYEDGALAAPSBEKPIPLAVPLILGSRKSVYKRLERFLR
1	H72607	ck: 1372	len: 387	1	hypothetical protein APE1326 - Aeropyrum
1					<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x(165)(L)x(S)(V)xx(L)(S)(R)xx(L)xxx(G)x(206) MAVSDSYKSLVDVHVKRLVAELRKGEEBFKRAASDMHARRRPIPCGNTITITGCGCTG LEPGASPEERIEFMGRVLVAGSALITLPIPTGTVDEADITLSPARRAKVNV VLGITLRVAEGILRLRASGAVENGELIERLRLPMWRGEOPLIYSIDLKRLSRKAREMGFVVLPAASCANVESHGCGCAACRUGP

E72626	ck: 6841	len: 146	1	hypothetical protein APE1469 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{100}(L)x(T)(V)xx(L)(S)(R)xx(M)xxx(A)x{30} MLSVAMDRFTLSLVPRFTTNGEFPALVLAIALANVADRIVFGWGRMLLEQWQCMWRRELOSS				
D72630	ck: 4651	len: 152	1	hypothetical protein APE1499 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{88}(A)x(P)(P)xx(L)(T)(R)xx(P)xxx(T)x{49} MSLRSSRSKSRMSRSGSTIVTALPGLIGMSLCSMTLLHLSAITSRTLSMPGMSGITPV				
H72538	ck: 9362	len: 175	1	hypothetical protein APE1597 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{39}(V)x(P)(V)xx(A)(C)(R)xx(V)xxx(G)x{120} MDECRHLVGYIYVLANMVYRVLAGSRHLAHRPHRDKPNVPACRTVYTMGAGTGDEVEYL				
D72545	ck: 2785	len: 194	1	hypothetical protein APE1648 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{16}(A)x(S)(M)xx(L)(T)(R)xx(A)xxx(V)x{162} MMGRGLKMTIGDRVLAIVSMGDLTKRYARAAVVPANSLMIMGGAAGALKRAGGSVTEEDAMRAH				
ALGVEGECPPADLARLV	1:				
E72557	ck: 8347	len: 132	1	hypothetical protein APE1745 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{36}(A)x(T)(V)xx(E)(S)(R)xx(A)xxx(L)x{80} MRVLPRGFSITLASSRVALLPESELILTSACRLNERAIVFFSKAPAPRTLPGMTISPLACIF				
H72557	ck: 8142	len: 181	1	hypothetical protein APE1748 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{94}(A)x(S)(I)xx(I)(S)(R)xx(P)xxx(T)x{71} MYFGSVTLGSPISVWLSCSRKAVLRSLILITSSSSSSSHRKNSTPRPAIILITASAMPITF				
TPCL	1:				
D72577	ck: 2318	len: 100	1	hypothetical protein APE1900 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{76}(P)x(S)(A)xx(A)(T)(R)xx(P)xxx(P)x{18} MIPVYVELNTSMRSEPFKSGTAMEPVLRSILSGMLNREMDLTITCSSIPSTLSFNNFSLKLS				
A72511	ck: 2835	len: 140	1	hypothetical protein APE2063 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{66}(L)x(T)(I)xx(L)(T)(R)xx(V)xxx(A)x{158} MLSSASMSDMPKTRLATSRASRLTSLTSLPSLNSLPKASVNTSAAGPLRSMLWLTFSHPSS				
C72529	ck: 9995	len: 310	1	hypothetical protein APE2207 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{45}(V)x(P)(V)xx(M)(S)(R)xx(L)xxx(A)x{249} MYRAEVRTESDIYDVAIVYFMSLAVRLVKTLEENRKTAGIQIVPYMYSRLLISTVAVLIG HVLDPSSVFEDELIGTAVKYGSDVIAHLEIRLNDIRLALARMELY LEIYLVAIVATIAIVFYIFFTSSVFGNIGVASQASTOLILYSFLVPLPINVIMLLVDRARPKEPIQIKDTW				
F72471	ck: 430	len: 133	1	hypothetical protein APE2415 - Aeropyrum	pe
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{82}(L)x(T)(I)xx(M)(T)(R)xx(M)xxx(S)x{35} MLVNRDISEMLPPTIATAPIESRPNTRATPVYAALTALEMLIAPRFPPTVAVAMLRKXADWTMO				
C70405	ck: 5289	len: 499	1	hypothetical protein aq_1220 - Aquifex	a
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{164}(A)x(S)(L)xx(L)(T)(R)xx(A)xxx(L)x{319} MMKRFELVLPVILILEFYNLGNAMNDESEFVDSAKNMKSGEFLTPYNGEIRLNKPPMT FVLGAVVFYLLRARELLKIKYVAGTIAVFLSGWMPFOYLVRHEEFLEVI KENKRIYALQOBPFYVYALDINVSFLPYSFFELWALKERRELSPLWFSFILFISYVKKIPVYIMPAYMAIITA				
D70454	ck: 3320	len: 269	1	hypothetical protein aq_1793 - Aquifex	a
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{4}(M)x(S)(I)xx(Y)(T)(H)xx(M)xxx(A)x{249} MMSPMASITMYTHFRMVKHALNWSSTLONFRTLAEILIFTLNLFILPYMNVFAFEIHN TEGTPRDFEFLLTITAFESKLTKRKIAPEUVEFSSPVGFSIFLFEQFMSHLF PPLFLITVAFSMTLITLIFGVISVSYLLMKRECK				
A75196	ck: 70	len: 147	1	hypothetical protein PAB0052 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{57}(M)x(P)(P)xx(A)(S)(R)xx(L)xxx(I)x{74} MSSLSLAMPPLNSISGATTLTSFVAFKAEFLITNSLVLSGKSPSSSLKPTMTSPPR				
E75204	ck: 8207	len: 179	1	hypothetical protein PAB0107 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{96}(M)x(T)(I)xx(A)(S)(H)xx(L)xxx(S)x{67} MDGYENKATNIKIERHIMKRDQDFLSSNPEISPLILVIGSIFPDLDVFTFFSESLALH				
H75185	ck: 6767	len: 266	1	hypothetical protein PAB3346 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{190}(I)x(P)(L)xx(I)(S)(R)xx(A)xxx(M)x{160} MRVATPRDIDDMVSVFLDAVNTGPRDSVAKSMETLSKVCQNGCIIAFIDGKPYMGCFILY VENEGCGMYRGRITPPLVAISRKAIRKAKELMAGRELIVPDANEDAMDILKRF SPVELTSCYRMRGLGKVEENHMYFGILITVAKG				
D71228	ck: 1741	len: 135	1	hypothetical protein PH0086 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{42}(I)x(P)(P)xx(A)(S)(R)xx(L)xxx(I)x{77} MGATTLTLEIVAISKALLTISTNSFVLSGKVSPLSKPTTISPPRARSREAAIILIKTFTL				
D71110	ck: 6672	len: 102	1	hypothetical protein PH0551 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{21}(P)x(P)(V)xx(A)(T)(R)xx(A)xxx(P)x{65} MLTILSLGNNIAEFNVVLPDEPVPAPATRMFAGLTTPRSTSORNAATSEFVYLVNFISIT				
E71120	ck: 8797	len: 104	1	hypothetical protein PH0732 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{42}(I)x(S)(L)xx(V)(S)(R)xx(L)xxx(L)x{46} MAHNHHYTPPKLHCHVAIPTEGNGHYIVVSPKSLSCYQIPISIRVYSKLSIYLLNSC				
H71124	ck: 5357	len: 272	1	hypothetical protein PH0767 - Pyrococcus	ccu
1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W x{43}(I)x(S)(L)xx(A)(T)(R)xx(L)xxx(L)x{213} MANENRLLSGSSAVRTPGSPALVMAITLLEPLGKCCENALMLSSIVFATKTPCLCNAASYI VIXVAFSTVFTIIPROGPMILIPDMNISSSSSEFSPSPSPAPAVIITPTL TPFSPHIVSVFTNLGTTIANKSSISGISTLTLVALTP				

[illegible]

HAADLVVITVTEFLHDKRASLLIGHGSDNSGDEKGLVTVANNKTLNLSRGRSA
SAQSTSTTTTRTCRMVSPHARVSCLEFTEFTFLSAPRRLLPFLRLMDLLDLTLVWARTLPURVL

1: MKRSVIEPSSDLKTLIAKAPVPSMIDVNRADVVRKASLEEDVATLNRPLKPSASLPSSRSK

1: D72485 ck: 7054 len: 242 1 probable high-affinity branched-chain am

1: MEETIILDRVDYKRRFGGIEALKVSFSVRGSEVGLIGNGAGKTTLENIISGTYMDRGRK

1: LEAVMGTCG

1: B71024 ck: 6905 len: 249 1 probable lipote protein ligase - Pyrocco

1: MRPLPLIVARPEOMADIALIARSSEGRVDOTRLYFKRSSTIGRFQSVAYDVNLKAKE

1: LKDGYSISIERATYTERGGINGVNREMEYVLRNFFAFPLDSDGELTEVLEIVE

1: KLIERIGNDKNFKOR

1: S72288 ck: 9584 len: 75 1 ribosomal protein L23 - Plasmodium falci

1: <(X){1,200}(L,I,V,M,A,P)x(P,T,S)(L,I,V,M,A,P)xx(L,I,V,M,A,F,Y,W

1: x(26)(I)x(S)(I)xx(F)(T)(K)xx(I)xxx(I)x(133)

1: MKEVLENYLNLIFYKINLVNFCILYISIKIFYTKLDIYIKINFIKTLINYNKININ

1: T10349 ck: 889 len: 374 1 very-late factor 1 protein - Orgyia pseu

1: <(X){1,200}(L,I,V,M,A,P)x(P,T,S)(L,I,V,M,A,P)xx(L,I,V,M,A,F,Y,W

1: MDGIQVNETTFENDMDARIQSAREHVFEDATROBCTPDEVNDSLMSKYMFPPFAPFTLL

1: DSKTKLNSDYVHDRGLIGALVFCIMJGTGRKINSEARLSDVDLNVLLIRKGLRS

1: DTILKRRRSRKNLTNNIKTKPLELAREIYANPVLQISKNSTPFRDRRLDDEAGVEMERPSNRHRLSLSNLYNSGVPLQ

1: S78187 ck: 3067 len: 197 1 ymf39 protein - Reclinomonas americana (

1: <(X){1,200}(L,I,V,M,A,P)x(P,T,S)(L,I,V,M,A,P)xx(L,I,V,M,A,F,Y,W

1: x(97)(I)x(S)(L)xx(F)(T)(K)xx(V)xxx(I)x(84)

1: MSLSIQKHYTLVLFVLAITLADFLIIVHDELIVLCITLFLTFILYVALKMYASNSDRRA

1: GIRLEGAISLDCKEANS

1: T11851 ck: 3584 len: 324 1 ruvB protein - Thermus aquaticus thermop

1: <(X){1,200}(L,I,V,M,A,P)x(P,T,S)(L,I,V,M,A,P)xx(L,I,V,M,A,F,Y,W

1: x(149)(L)x(T)(A)xx(L)(S)(R)xx(I)xxx(L)x(159)

1: VEDIALKPTLDYIGDERLQKLVSEYLEAKAREPLEHLLGFPGLCKTTLAHVIAHELG

1: MRDARLLGVITEALEIGRSGRMVARKLFFRVADQVGEVEVITERRALE

1: ALALAGIDELGELERKRDREILEVALILRFAGAPGLATATVALSDEDPGLEVEHEPYLIRGOGLKRTPRGAVATELAYRHLGYPPPV

1: JCS705 ck: 3534 len: 125 1 alpha-amylase inhibitor 0.19 - synthetic

1: <(X){1,200}(L,I,V,M,A,P)x(P,T,S)(L,I,V,M,A,P)xx(L,I,V,M,A,F,Y,W

1: x(92)(A)x(S)(I)x(S)(V)(C)(R)x(I)xxx(A)x(117)

1: MSGPMPCYPGAGFVPLPACRPLRLTRCNGSGVPEAVVLRDCCOQLAHISEMCRGALYS

Databases searched:

NRRF, Release 62.0, Released on 30sep1999, Formatted on 26oct1999

Total lengths: 561

Total sequences: 47,169,319

101:CCGCF2T - cytochrome c2 - Rhodocyclus tenuis

CPD time: 03:25.06

11:AA_SEQUENCE 1.0

C:Species: Rhodocyclus tenuis
C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 15-Jan-1999
C:Accession: A00090
R:Ambler, R.P.; Meyer, T.E.; Kamen, M.D.
A:Title: Anomalies in amino acid sequences of small cytochromes c and cytochromes c' from two species of purple photosynthetic bacteria.
A:Reference number: A93207; MUID:79199668
A:Accession: A00090
A:Molecule type: protein
A:Residues: 1-92 <AMB>
C:Comment: This sequence is more closely related to the sequences of cytochrome c551 from Pseudomonas and Azotobacter than to the sequences of cytochrome c2 from other species of Rhodospirillum.
C:Superfamily: cytochrome c6; cytochrome c6 homolog
C:Keywords: chromoprotein; electron transfer; heme; iron; photosynthesis
F:1-83/Domain: cytochrome c6 homology <CYC>
F:12,15/Binding site: heme (Cys) (covalent) #status predicted
F:16,66/Binding site: heme iron (His, Met) (axial ligands) #status predicted
CCQFZT Length: 92 February 11, 2000 15:51 Type: P Check: 2620 ..

1 ADESLAQTK GCLACHNEK KVGPAVGV AKKVGAGGA EARLVAKYNA
51 GGGVAKOL GAELPMANN VKKATRLV KWLISLKQID YK

!!A:SEQUENCE 1.0
P1:RDHCB - carbonyl reductase (NADPH) (EC 1.1.1.184) - human
N:Alternate names: aldehyde reductase I; NADPH-dependent carbonyl reductase; prostaglandin G-ketoreductase; xenobiotic ketone reductase
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 11-Jun-1999
C:Accession: B34340
R:Peoples, O.P.; Sinskey, A.J.
J: Biol. Chem. 264, 15293-15297, 1989
A:Title: Poly-beta-hydroxybutyrate biosynthesis in Alcaligenes eutrophus H16. Characterization of the genes encoding beta-ketothiolase and acetoacetyl-CoA reductase.
A:Reference number: A34340; MUID:89359356
A:Accession: B34340
A:Molecule type: DNA
A:Residues: 1-246 <PEO>
A:Cross-references: EMBL:J04987; NID:9141953; PIDN:AAA21973.1; PID:9141955
C:Keywords: NAD; oxidoreductase; poly-beta-hydroxybutyrate biosynthesis
F:4-184/Domain: short-chain alcohol dehydrogenase homology <SADH>

RDALAE Length: 246 February 11, 2000 15:51 Type: P Check: 2829 ..

1 MGRIRAVTG GNGGIGTALC QRLAKDFRY VAGCCPSNPR REKLEQDKA
51 LGDFLASEG NVADWDSTKT AFDKXSEVG EVDVLIINAG ITRDVFRRM
101 TRADWDVAID TNLISLENT KOVIDGADR GMGRIVNIS VNGOKGQFG
151 TNSYAKAGL HGTMALAOE VARKGVYNT VSPGIADIM VKAIRQDVID
201 KIYATIPYK LGPEELASI CAMLSSESG PSTGADFSLN GGLHMG

!!A:SEQUENCE 1.0
P1:SDS397 - granatoin polyketide ketoreductase (EC 1.1.1.-) graIII - Streptomyces violaceoniger
N:Alternate names: 3-hydroxyacyl-CoA dehydrogenase homolog 1
C:Species: Streptomyces violaceoniger
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S05397
R:Sherman, D.H.; Malpartida, F.; Bibb, M.J.; Kleser, H.M.; Bibb, M.J.; Hopwood, D.A.
EMBO J. 8, 2717-2725, 1989

variety of carbonyl compounds to the corresponding alcohols.
C:Genetic: GDB:GDB
A:Gene: GDB:GDB
A:Cross-references: GDB:126610; OMIM:114830
A:Map position: 21q22.12-21q22.12
A:Introns: 97/1; 133/1
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: blocked amino end; monomer; NADP; oxidoreductase
F:6-185/Domain: short-chain alcohol dehydrogenase homology <SADH>
F:239/Modified site: N6-1-carboxyethyllysine (Lys) (partial) #status experimental

RDHCB Length: 277 February 11, 2000 15:51 Type: P Check: 7216 ..

1 MSSGHVALV TGSNGIGLA IVRDLRLEF GDVVLITARDV TRGAAVQOL
51 QAEGLSPRF QLDIDLOSI RALRDLRKE YGGLDVLVNN AGIAFKVADP
101 TPFHCAEVT MKNFPGTBD VCTLLPLIK PGRVNVASS IMSVRLKSC
151 SPELOQKFRS ERTTEELVG LMKFVEDTK KGVHKEGWP SSAGVTRIG
201 VTVLSRIHAR KLSQRKDK ILNACCPGW VRTDMAGPRA TKSPEGAET
251 PVYLLALPP AEGPHQGEVS EKREVEOW

A:Title: Structure and deduced function of the granatidin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tue22.
 A:Reference number: S05393; MUID:90060034
 A:Accession: S05393
 A:Residues: 1-372 <SMH>
 A:Molecule type: DNA
 A:Cross-references: EMBL:X16144; NID:947976; PIDN:CAA34263.1; PID:947978
 C:Genetics:
 C:Gene: gratii
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: NAD; oxidoreductase
 F:18-199/Domain: short-chain alcohol dehydrogenase homology <SADH>
 S05397 Length: 272 February 11, 2000 15:51 Type: P Check: 3104 ..

1 MTATATATA TPTAAKFA IYTGATSGIG LAIRRLAL GARTLCARD
 51 EERLQTVKE LRGEFVDG TVCDVADPAQ IRAYVAAVQ RYGVDIIVN
 101 NAGRGGGAT AEIADLMID VITNTISVF LMRKVNAG GMLAKRGRI
 151 INIATGGKQ GYVAHPYSA SRGVVGLTK ALGELARTG ITVNAVCPGF
 201 VETPAERVR EHYAGIQVS EETEDRITN RVLGRVET REYAAVEXL
 251 VADDAAVTA QALNVCGLG NY

11AA_SEQUENCE 1.0
 P1:S2079 - monensin polyketide ketoreductase (EC 1.1.1.-) - *Streptomyces cinamonensis*
 C:Species: *Streptomyces cinamonensis*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S25079; S18170
 R:Arrowsmith, T.J.; Walpertida, F.; Sherman, D.H.; Birch, A.; Hopwood, D.A.; Robinson, J.A.
 Mol. Gen. Genet. 234, 254-264, 1992
 A:Title: Characterisation of acti-homologous DNA encoding polyketide synthase genes from the monensin producer *Streptomyces cinamonensis*.
 A:Reference number: S25076; MUID:92374994
 A:Accession: S25079
 A:Molecule type: DNA
 A:Residues: 1-261 <ARR>
 A:Cross-references: EMBL:Z11511; NID:946799; PIDN:CAA77599.1; PID:946803
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: NAD; oxidoreductase
 F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>
 S25079 Length: 261 February 11, 2000 15:51 Type: P Check: 4473 ..

1 MGTSTSRVAL VTGATSGIGL ATRRLAAG HLVEFGARTE SDYATVYKAL
 51 RNDGLEAEQ VLDVVDGASV TAFVQNAVDR YGRIDVLVNN AGRSGGCVTA
 101 DLTLDELMDV IDTNLSVFR MTRAVLTGG MTRERGRIT NVASTGGKQ
 151 VILGAPYSAS KHGVVGFKA LGNELAPGI TVNAVCPGV ETMAQRVRO
 201 GYAAVDTTE EAILTKFOAK IPLGRSTPE EVAGLIGTILA SDTAASITSO
 251 ALNVCGLGN F

11AA_SEQUENCE 1.0
 P1:B42147 - 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) - *Escherichia coli*
 N:Alternate names: 3-ketoacyl-ACP reductase
 C:Species: *Escherichia coli*
 C:Date: 30-Sep-1993 #sequence_revision 31-Oct-1997 #text_change 11-Jun-1999
 C:Accession: B64853; B42147; C41856
 R:Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Berna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Lasker, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617
 A:Accession: B64853
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-244 <BLAT>
 A:Cross-references: GB:AE000210; GB:U00096; NID:91787332; PIDN:AACT41177.1; PID:91787335; UMG:U0093
 A:Experimental source: strain K-12, substrain MG1655
 R:Rawlings, M.; Cronan Jr., J.E.
 J. Biol. Chem. 267, 5751-5754, 1992
 A:Title: The gene encoding *Escherichia coli* acyl carrier protein lies within a cluster of fatty acid biosynthetic genes.
 A:Reference number: A42147; MUID:92210530
 A:Accession: B42147
 A:Molecule type: DNA
 A:Residues: 1-29, 6', 31-244 <RAM>
 A:Cross-references: GB:M67040; NID:9145879; PIDN:AAA23739.1; PID:9145881
 R:Verwoert, I.I.; Verdree, E.C.; van der Linden, K.H.; Nijkamp, H.J.; Stultje, A.R.
 J. Bacteriol. 174, 2851-2857, 1992
 A:Title: Cloning, nucleotide sequence, and expression of the *Escherichia coli* fabD gene, encoding malonyl coenzyme A-acyl carrier protein transacylase.
 A:Reference number: A41856; MUID:92234941
 A:Accession: C41856
 A:Molecule type: DNA
 A:Residues: 1-42, R', 44-45 <VER>
 A:Cross-references: GB:M67040; NID:9145885; PIDN:AAA23743.1; PID:9145888
 A:Note: sequence inconsistent with the nucleotide translation (NCBI:97150, NCBI:97150)
 C:Genetics:
 A:Gene: fabG
 A:Map position: 24 min
 C:Function: fatty acid biosynthesis
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: fatty acid biosynthesis; NADP; oxidoreductase
 F:6-18/Domain: short-chain alcohol dehydrogenase homology <SADH>
 F:6-36/Region: beta-alpha-beta motif; nucleotide binding fold
 F:151/Active site: Tyr #status Predicted
 B42147 Length: 244 February 11, 2000 15:51 Type: P Check: 3492 ..

1 MNEGRKIALV TGASRGITRA IAEFLAARGA KVIGRTSEN GAQISDYIG
 51 ANKGMLANV IDPRASIEVL EKRAEFGV DILVNNAGIT RDLNLRMD
 101 EEMNDILLEN LSVPRLSKA VMRAKKRRH GRITIGSV GINGNGGOAN
 151 YAAKAGLIG FSKSLAREVA SRGITVYVA PGFIETDMTR ALSDDRAGI
 201 LAQVPAGRLG GAQELANVA FLASDEAVI TGETLHVNG MYWV

11AA_SEQUENCE 1.0
 P1:D69930 - probable 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) - *Bacillus subtilis*
 C:Species: *Bacillus subtilis*
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: D69930; S01270
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Bertero, M.G.; Bessieret, P.; Bolotin, A.; Borcherdt, S.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, J.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Dusterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, B.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Gutseppl, G.; Guy, B.J.; Haga, K.; Hahle, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iitaya, M.; Itoya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kraer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle, D.; Potwilk, S.; Prescott, A.M.; Pressac, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roegner, B.; Rose, M.; Sadate, Y.; Sato, T.; Scallion, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takegi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassaret, A.; Viari, A.; Wamburt, R.; Wedler, E.; Wedler, H.; Weisenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: D69930

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1238 <KUN>

A:Cross-references: GB:289114; GB:AL009126; NID:92634230; PIDN:CA13743.1; PID:EL15323; PID:92634244

A:Experimental source: strain 168

R:Cartigan, C.M.; Haasman, J.A.; Smith, M.T.; Wake, R.G.

Nucleic Acids Res. 15, 8501-8509, 1987

A:Title: Sequence features of the replication terminus of the *Bacillus subtilis* chromosome.

A:Reference number: S01270; MUID:98040469

A:Accession: S01270

A:Molecule type: DNA

A:Residues: 62-338 <CAR>

A:Cross-references: EMBL:X06168; NID:940205; PIDN:CA29533.1; PID:9809662

C:Genetics:

A:Gene: yoxD

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:7186/Domain: short-chain alcohol dehydrogenase homology <SADH>

D69930 Length: 238 February 11, 2000 15:51 Type: P Check: 2171

1 MOSLOKHTAL ITGGGGRIGR ATALALAKES VNIGLIGRTS ANEKVAEEV

51 KALGVAAARA AADYKADAOV NQAVAYKEQ LGDIDILNN AGSKFGGFL

101 DLSADEMENT IOYVLMGYTH VTRAVPEMI ERKADILNI SSTAQRGAA

151 VTSAYSAKRF AVLGITBSLM QEVKRNHIV SALTSTYAS DMSLEMLND

201 GNPEKYWQPE DIAETVAOL KLDPRITIKT AGLWSTNP

11AA-SEQUENCE 1.0

P1:DECEP - N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) - *Escherichia coli*

C:Species: *Escherichia coli*

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 11-Jun-1999

C:Accession: J03032; A42377; A65203; A50776

R:Parrot, C.; Boyen, A.; Cohen, G.N.; Glansdorff, N.

Gene 68, 275-283, 1988

A:Title: Nucleotide sequence of *Escherichia coli* argB and argC genes: comparison of N-acetylglutamate kinase and N-acetylglutamate-semialdehyde dehydrogenase with homologous and analogous enzymes.

A:Reference number: J03031; MUID:89121510

A:Accession: J03032

A:Molecule type: DNA

A:Residues: 1-334 <PAR>

A:Cross-references: GB:N21446; NID:9145332; PIDN:AA23477.1; PID:9145333

R:Meinert, T.; Schmitt, E.; Mechlam, Y.; Blanquet, S.

J. Bacteriol. 174, 2323-2331, 1992

A:Title: Structural and biochemical characterization of the *Escherichia coli* argB gene product.

A:Reference number: A42377; MUID:92202162

A:Accession: A42377

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-19 <MEP>

A:Cross-references: GB:X55417

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A65203

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-334 <BLAT>

A:Cross-references: GB:AE000470; GB:U00096; NID:92367332; PIDN:AACT6940.1; PID:91790396; UMG:93938

A:Experimental source: strain K-12, substrain MG1655

C:Comment: In arginine biosynthesis glutamate is first converted to N-acetylglutamate, which leads to N-acetylglutamate phosphate and then to acetylglutamate kinase and N-acetyl-gamma-glutamyl-phosphate reductase.

C:Genetics:

A:Gene: argC

A:Map position: 90 min

C:Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase

C:Keywords: arginine biosynthesis; oxidoreductase

F:154/Active site: Cys #status predicted

DECEP Length: 334 February 11, 2000 15:51 Type: P Check: 6342

1 MENTLVGAS GYAGAEVTV VNRHPHMT ALTVASOSND AGKLSDHP

51 QKGYDLEPL QPMDSIEFS PGVDVFLAT AHEVSHDAP QLEAGCVF

101 DLSGARVND ATEFYKYYG THOYELLEQ AAVGLAEMCG NTKRANLIA

151 VPCCTPAAQ LALRPILDAD LLDINQWPI NATSGVSGA RRAAISNSPC

201 EYSLQPYGF THROPELAT HIGADVITP HLGFPFGIL EITTCRLSG

251 VTGAQVAQVL QQAVAHKPLV RLYDKGVAL KNYVGLPFCF IGFVAOGEHL

301 IVATEDNLL KGAQAQAVOC ANIRFGYAEI QSLI

11AA-SEQUENCE 1.0

P1:DECC - aspartate carboxyltransferase (EC 2.1.1.32) catalytic chain - *Escherichia coli* (strain K-12)

N:Alternate names: aspartate transcarboxylase catalytic chain; aspartyl carbamoyltransferase catalytic chain; carbamylaspartotransferase catalytic chain

C:Species: *Escherichia coli*

C:Date: 30-Nov-1980 #sequence_revision 10-Oct-1997 #text_change 11-Jun-1999

C:Accession: H65236; A00561; A21121; B36599; A21120; S56471

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: H65236

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-311 <BLAT>

A:Cross-references: GB:AE000495; GB:U00096; NID:92367361; PIDN:AACT7202.1; PID:92367364; UMG:94245

A:Experimental source: strain K-12, substrain MG1655

R:Schachman, H.K.; Paua, C.D.; Navre, M.; Karelis, M.J.; Wu, L.; Yang, Y.R.

Proc. Natl. Acad. Sci. U.S.A. 81, 115-119, 1984

A:Title: Location of amino acid alterations in mutants of aspartate transcarboxylase: structural aspects of interallelic complementation.

A:Reference number: A93985; MUID:84119419

A:Accession: A00561

A:Molecule type: DNA

A:Residues: 1-149, 'E', 151-311 <SCH>
 A:Cross-references: GB:K01472; NID:9147463; PIDN:AAA24476.1; PID:9147464
 R:Hoover, T.A.; Roof, W.D.; Foltermann, K.F.; O'Donovan, G.A.; Bencini, D.A.; Wild, J.R.
 Proc. Natl. Acad. Sci. U.S.A. 80, 2462-2466, 1983
 A:Title: Nucleotide sequence of the structural gene (pyrB) that encodes the catalytic polypeptide of aspartate transcarbamoylase of *Escherichia coli*.
 A:Reference number: A21121; MUID:83195078
 A:Accession: A21121
 A:Molecule type: DNA
 A:Residues: 1-60, 'Q', 62-165, 'T', 167-220, 'V', 222-311 <HOON>
 A:Cross-references: GB:V00323
 R:Donahue, J.P.; Turnbough Jr., C.L.
 J. Biol. Chem. 265, 19091-19099, 1990
 A:Title: Characterization of transcriptional initiation from promoters P-1 and P-2 of the pyrB operon of *Escherichia coli* K12.
 A:Reference number: A35599; MUID:91035438
 A:Accession: B36599
 A:Molecule type: DNA
 A:Residues: 1-18 <DON>
 A:Cross-references: GB:M60508; NID:9147469; PIDN:AAA24481.1; PID:9147471
 R:Konysberg, W.H.; Henderson, L.
 Proc. Natl. Acad. Sci. U.S.A. 80, 2467-2471, 1983
 A:Title: Amino acid sequence of the catalytic subunit of aspartate transcarbamoylase from *Escherichia coli*.
 A:Reference number: A21120; MUID:83195079
 A:Accession: A21120
 A:Molecule type: protein
 A:Residues: 2-60, 'Q', 62-86, 'Q', 88-90, 'N', 92-129, 'N', 131-220, 'V', 222-256, 'D', 258-259, 'M', 261, 'A', 263-311 <KON>
 R:Ke, H.; Honzato, R.B.; Lipscomb, W.N.
 Proc. Natl. Acad. Sci. U.S.A. 81, 4037-4040, 1984
 A:Title: Structure of unliganded aspartate carbamoyltransferase of *Escherichia coli* at 2.6-angstroms resolution.
 A:Reference number: A93993; MUID:84248054
 A:Contents: annotation; X-ray crystallography, 2.6 angstroms; quaternary structure
 R:Burkett, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes.
 A:Reference number: S56314; MUID:95334362
 A:Accession: S56471
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-195, 'R', 197-311 <BUR>
 A:Cross-references: EMBL:U14003; NID:91263172; PIDN:AA97142.1; PID:9537087
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.
 C:Genetics:
 A:Gene: pyrB
 A:Map position: 97 min
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase homology
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transfease
 F:8-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
 DECC Length: 311 February 11, 2000 15:51 Type: P Check: 4302 ..
 1 MAPLXQKH ISINDSRD LNLVLAATAK LKANPPELL KHKVASCFF
 51 EASTRRRLSF ETSMRLGAS VVGFSDSANT SLGKKGELTA DTISIVTYY
 101 DAIVMHPDE GAARLATEFS GNVPLVNA GD GSNQHPQTQL LDLFTIETQ
 151 GRDNLHIM VGDLYKGRV HSLQALAKF DGNREFFIAP DALAMPQYI
 201 DMUDEGIAW SLHSIEEVM AEVDILYMT VQKERLPSE YANVKAQFV
 251 RASDLNANA NMKVLPPLR VDEIATDVOK TPNAVFOOA GNGIFAQAL
 301 LALVNRDYL L
 I:AA_SEQUENCE 1.0
 P:OWSBAC - aspartate carbamoyltransferase (EC 2.1.3.2) catalytic chain - *Salmonella typhimurium*
 N:Alternate names: aspartate transcarbamylase catalytic chain; carbamylaspartotranskinase
 C:Species: *Salmonella typhimurium*
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
 A:Accession: S00049
 R:Michaels, G.; Kelln, R.A.; Nargang, F.E.
 Eur. J. Biochem. 166, 55-61, 1987
 A:Title: Cloning, nucleotide sequence and expression of the pyrB1 operon of *Salmonella typhimurium* LT2.
 A:Reference number: S00028; MUID:87246692
 A:Accession: S00049
 A:Molecule type: DNA
 A:Residues: 1-311 <MIC>
 A:Cross-references: GB:X05641; NID:947861; PIDN:CAA29129.1; PID:947863
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.
 C:Genetics:
 A:Gene: pyrB
 A:Map position: 98 min
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase homology
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transfease
 F:8-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
 OMBAC Length: 311 February 11, 2000 15:51 Type: P Check: 4073 ..
 1 MAPLXQKH ISINDSRD LNLVLAATAK LKANPPELL KHKVASCFF
 51 EASTRRRLSF ETSMRLGAS VVGFSDSANT SLGKKGELTA DTISIVTYY
 101 DAIVMHPDE GAARLATEFS GNVPLVNA GD GSNQHPQTQL LDLFTIETQ
 151 GRDNLHIM VGDLYKGRV HPAKPTLAK ESGNREFFIYA PDALAMPQYI
 201 LMDLDEGMA WSLGSGIEV MADVDILYMT RVQKERLPDS EYANVKAQFV
 251 LRPDNGARE NMKVLPPLR IDEITDVOK TPNAVFOOA GNGIFAQAL
 301 LALVNSELS L
 I:AA_SEQUENCE 1.0
 P:OWSBAC - aspartate carbamoyltransferase (EC 2.1.3.2) catalytic chain - *Serratia marcescens*
 N:Alternate names: aspartate transcarbamylase catalytic chain; carbamylaspartotranskinase
 C:Species: *Serratia marcescens*
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
 A:Accession: B34396
 R:Beck, D.; Kedzie, K.W.; Wild, J.R.
 J. Biol. Chem. 264, 16629-16637, 1989
 A:Title: Comparison of the aspartate transcarbamoylases from *Serratia marcescens* and *Escherichia coli*.
 A:Reference number: A34396; MUID:89380286
 A:Accession: B34396
 A:Molecule type: DNA
 A:Residues: 1-306 <BECK>
 A:Cross-references: GB:J05033; NID:9398074; PIDN:AA26564.1; PID:9398075
 C:Comment: The active enzyme contains two trimers of catalytic chains and three dimers of regulatory chains; it catalyzes the rate-limiting step in the biosynthesis of pyrimidines.
 C:Genetics:
 A:Gene: pyrB
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine

carbamoyltransferase homology
 C:Keywords: heterododecamer; homohexamer; homotrimer; pyrimidine nucleotide biosynthesis; transference
 F:8-300/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
 OMSAC Length: 306 February 11, 2000 15:51 Type: P Check: 6543 ..

1 MANPLYRKH IISINDLRD LELVLAIPA LKANPOPELL KHYVASCFF
 51 EASTRRLSE ETSMRLGAS VVGFDGSGNT SLCKKGETIA DTISVSTIV
 101 DAIVMRHPOE GARMASEFSG NNPVLNACDG NQHPOTQLD LFTIETQGR
 151 LSNLSIAVWG DKYGRVHS LQALAKREG NREYFIAPDA LAMPAYILKM
 201 LEERKEEYS HGSEIEVPE LDILMTVQ KERLDPESEA NKKAQFVLAA
 251 DIAGANLKV LHPRLRDEI AIDVAKTPHA YFQQAQNGI FARSLALIV
 301 NADLAL

11AA:SEQUENCE 1.0
 P1:OWP - ornithine carbamoyltransferase (EC 2.1.3.3) - fission yeast (Schizosaccharomyces pombe)
 N:Alternate names: citrulline phosphorylase; ornithine transcarbamylase
 C:Species: Schizosaccharomyces pombe
 C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 11-Jun-1999
 C:Accession: S22390; S22175
 Eur. J. Biochem. 205, 33-43, 1992
 Rivan Hufel, C.; Dubois, E.; Messenguy, F.
 A:Title: Cloning and sequencing of arg3 and arg1 genes of Schizosaccharomyces pombe on a 10-kb DNA fragment. Heterologous expression and mitochondrial targeting of their translation products.
 A:Reference number: S22389; MUID:92209520
 A:Accession: S22390
 A:Molecule type: DNA
 A:Residues: 1-327 <V>
 A:Cross-references: EMBL:X63577; NID:94907; PIDN:CAA45133.1; PID:94908
 C:Genetics: arg3
 A:Gene: arg3
 C:Map position: 11
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase homology
 C:Keywords: arginine biosynthesis; mitochondrion; transference
 F:8-315/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
 OWP Length: 327 February 11, 2000 15:51 Type: P Check: 9240 ..

1 MSRRKPRHL LSRDLRGE IVKLIDRSE IKQAKKQFO NRRSQMSGL
 51 SSONVAMITS KRSTRVSV ESAVSCLGN AMFLKRDIO LGVNESLYOT
 101 SKVSSKWSG IYARVNTYSD VATLAKHASC PYINGLCDF HPLOALADLL
 151 TIKETKSPD GLKVAWGA NVVLIDMIA NAKVGIHVS AKRPVNVKD
 201 DILSIWEMA NENGSTFEIV NDKPVAVKNA DIVVTDTWIS MGOAEKEQR
 251 LKQFTGQVY GEIMKLKPS CKMECLPRH PEEVSDEVYF GENSLVFOEA
 301 ENRKWTIVAV LPAALVNRGE ILPPASA

11AA:SEQUENCE 1.0
 P1:A69962 - phosphate butyryltransferase (EC 2.3.1.19) yqis - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A69962
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Bester, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boudisier, L.; Brans, A.; Braun, M.; Bignelli, S.C.; Bron, S.; Brouillette, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, S.K.; Codani, J.-C.; Conerton, J.F.; Cummings, N.T.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Ertington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fittz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizel, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Hada, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kaashara, Y.; Klierer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lander, T.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maude, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noack, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porrolik, S.; Prescott, A.M.; Prescan, E.; Pujo, P.; Fumelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sedate, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Solito, B.; Sorokin, A.; Taccou, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, R.; Vanlier, F.; Vassarotti, A.; Viaril, A.; Wambull, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: A69962
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-299 <KUN>
 A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CAB14340.1; PID:9185677; PID:92634843
 A:Experimental source: strain 168
 C:Genetics: yqis
 A:Gene: yqis
 C:Superfamily: phosphate acetyltransferase
 C:Keywords: acyltransferase
 A69962 Length: 299 February 11, 2000 15:51 Type: P Check: 158 ..

1 MKIKDIGRA SIHKNTIIV AHADEEVIR AVKLAHEHS ARFLITGDSK
 51 KINELTSMQ GHVEIYVHAN TPEESKIAV RAVHKTADV LMKGVNPIV
 101 LKAVLNROE GLRSAYLSH VAVFDIPED RLMEFTDSAM NIAPSELEL
 151 QILONAVHA HAVGNMPPKA AALAAEVTVN PKMEATVNA ALAOMYKRGQ
 201 IKGCIYDPL ALDNAVSOIA AAOKKISGDV AGNDILLV TIEAGNIIYK
 251 SLIFAKASV AAVITGAKAP IALTSRADSA ENKLYSIALA ICASEEYTH

11AA:SEQUENCE 1.0
 P1:XNEBHC - histidinol-phosphate transaminase (EC 2.6.1.9) - Salmonella typhimurium
 N:Alternate names: histidinol-phosphate aminotransferase
 C:Species: Salmonella typhimurium
 C>Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 28-May-1999
 C:Accession: U00158
 R:Carlomagno, M.S.; Chiarotoli, L.; Alfano, P.; Napo, A.G.; Bruni, C.B.; Mol, B.O.L. 203, 585-606, 1988
 A:Title: Structure and function of the Salmonella typhimurium and Escherichia coli K-12 histidine operons
 A:Reference number: J00131; MUID:89094629
 A:Accession: U00158
 A:Molecule type: DNA
 A:Residues: 1-359 <AR>
 A:Cross-references: GB:X13464; NID:947719; PIDN:CAA31824.1; PID:9477723
 C:Comment: This enzyme catalyzes the conversion of imidazolephosphate to L-histidinol phosphate, the eighth step in histidine biosynthesis.
 C:Genetics: hisC
 A:Gene: hisC

A:Map position: 42 min
C:Superfamily: histidinol-phosphate aminotransferase
C:Keywords: aminotransferase; histidine biosynthesis

XNBBHC length: 359 February 11, 2000 15:51 Type: P Check: 1849 ..

1 MSTEINTLSVA DLARENVRNL VPYOSARRLG GNGDVNLNAN EEPFAVEFOL
51 TOOTLNRYPE COPKAVIENY AQYAGVPRQ VLVSRGADRG IELVIRARCE
101 PGKRALIYCP PTYGMYSVA ETIGVERTV PALENWOLD QGISNDLDT
151 KVFVFCSPNN PTGOLINPDD LRTLELLELRG KALVADEAV IEFQCAATLI
201 GMLVEYPHLV IRLTSLKAFK LAGLRGFTL ANEEVNLNL KVIAEYPIST
251 PVADIAAQS CPOESNAMD RVAOTVDERQ YLVNALQDTA CVENHFDSET
301 NYLIARFTAS SSYFSLWMD GILRDONKO PSLSGCLRTIT VGTROENOPV
351 IDALRAEPV

!AA_SEQUENCE 1.0
P1:S29090 - dual specificity phosphoprotein phosphatase (EC 3.1.3.-) 1 - human
N:Alternate names: protein-tyrosine-phosphatase Cl100;
protein-tyrosine-phosphatase, nonreceptor type 10
C:Species: Homo sapiens (man)
C>Date: 25-Feb-1994 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: S29090; A53052
R:Keyes, S.M.; Emslie, E.A.
Nature 359, 644-647, 1992
A:Title: Oxidative stress and heat shock induce a human gene encoding a
protein-tyrosine phosphatase.
A:Reference number: S29090; MUID:93024952
A:Accession: S29090
A:Molecule type: mRNA
A:Residues: 1-367 <KEY>
A:Cross-references: EMBL:X68277; NID:g29980; PID:CAA48338.1; PID:g29981
R:Kwak, S.P.; Hakes, D.J.; Matteil, K.D.; Dixon, J.E.
J. Biol. Chem. 269, 3596-3604, 1994
A:Title: Isolation and characterization of a human dual specificity
protein-tyrosine phosphatase gene.
A:Reference number: A53052; MUID:94148864
A:Accession: A53052
A:Molecule type: DNA
A:Residues: 1-367 <KWA>
A:Note: sequence extracted from NCBI backbone (NCBIN:143800, NCBIN:143802,
NCBIN:143804, NCBIN:143806, NCBI:143807)
C:Genetics:
A:Gene: GDB:DUSP1; PTN10
A:Cross-references: GDB:136197; OMIM:600714
A:Map position: 5q34-5q34
C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual
specificity phosphoprotein phosphatase homology
C:Keywords: heat shock; phosphoprotein; phosphoric monoester hydrolase;
stress-induced protein
F:101-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology
<VHL>
F:228/Active site: Cys (phosphocysteine intermediate) #status predicted
F:264/Binding site: substrate phosphate (Arg) #status predicted

S29090 length: 367 February 11, 2000 15:51 Type: P Check: 5401 ..

1 MMEVEGILDA GGLRALRGER AAOCLLDCCR SFPAFAGHI AGSVNVRST
51 IYRRRAKGM GLEHIVPNAE LRGRLLAGAY HAVVILDEKS AALDCAKRDG
101 TLALAGALC REARAQVFF LKGYEAFSA SCEPLCSKOS TPMGLSLPLS
151 TSPVDSAESG CSSCSTPLYD OGGVEILPF LVYGSAYHVS RKDMDALGI
201 TALINVSANC PNHFEHGYO KSIPVEDNHK ADLSSFNENA IDFIDSINA

251 GGRVYFHQA GISRATITCL AYLMRTNRK LDEAFEPKO RRSIISNFS
301 FMGOLLQFES OYLAPHCAS AGSPANAVLD RGTSTTVFN FPVSPVHST
351 NSALSTYQSP ITTSPSC

!AA_SEQUENCE 1.0
P1:S24411 - dual specificity phosphoprotein phosphatase (EC 3.1.3.-) 1 - mouse
N:Alternate names: 3CH134 protein; protein-tyrosine-phosphatase erp,
nonreceptor type 10
C:Species: Mus musculus (house mouse)
C>Date: 19-Feb-1994 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A54681; S24411
R:Noguichi, T.; Metz, R.; Chen, L.; Matteil, M.G.; Carrasco, D.; Bravo, R.
Mol Cell Biol 13, 5195-5205, 1993
A:Title: Structure, mapping, and expression of erp, a growth factor-inducible
gene encoding a nontransmembrane protein tyrosine phosphatase, and effect of
erp on cell growth.
A:Reference number: A54681; MUID:93360956
A:Accession: A54681
A:Molecule type: DNA
A:Residues: 1-367 <MOG>
A:Cross-references: GB:S64851; NID:g409976; PIDN:AAB27882.1; PID:g409977
R:Charles, C.H.; Abler, A.S.; Lau, L.F.
Oncogene 7, 187-190, 1992
A:Title: cDNA sequence of a growth factor-inducible immediate early gene and
characterization of its encoded protein.
A:Reference number: S24411; MUID:92158357
A:Accession: S24411
A:Molecule type: mRNA
A:Residues: 1-367 <CHAS>
A:Cross-references: EMBL:X61940; NID:g49735; PIDN:CAA43944.1; PID:g49736
C:Genetics:
A:Gene: erp
A:Introns: 123/1; 172/1; 245/1
C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual
specificity phosphoprotein phosphatase homology
C:Keywords: immediate-early protein; phosphoprotein; phosphoric monoester
hydrolase
F:101-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology
<VHL>
F:238/Active site: Cys (phosphocysteine intermediate) #status predicted
F:264/Binding site: substrate phosphate (Arg) #status predicted

S24411 length: 367 February 11, 2000 15:51 Type: P Check: 5871 ..

1 MMEVEGILDA GGLRALRGER AAOCLLDCCR SFPAFAGHI AGSVNVRST
51 IYRRRAKGM GLEHIVPNAE LRGRLLAGAY HAVVILDEKS ASLDCAKRDG
101 TLALAGALC REARSTQVFF LKGYEAFSA SCEPLCSKOS TPTGLSLPLS
151 TSPVDSAESG CSSCSTPLYD OGGVEILSF LVYGSAYHVS RKDMDALGI
201 TALINVSANC PNHFEHGYO KSIPVEDNHK ADLSSFNENA IDFIDSINKA
251 GGRVYFHQA GISRATITCL AYLMRTNRK LDEAFEPKO RRSIISNFS
301 FMGOLLQFES OYLAPHCAS AGSPANAVLD RGTSTTVFN FPVSPVHPT
351 NSALNTKSP ITTSPSC

!AA_SEQUENCE 1.0
P1:PRMSCL - cytotoxic T-lymphocyte proteinase (EC 3.4.21.-) 1 precursor - mouse
N:Alternate names: CTLA-1 protein; cytotoxic cell proteinase 1 (CPT1);
cytotoxic T-cell-specific proteinase 1; cytotoxic T-cell-specific proteinase
granzyme B
C:Species: Mus musculus (house mouse)
C>Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 18-Jun-1999
C:Accession: A94286; A93382; A28952; B26944; I48937; A00956
R:Loeb, C.G.; Finlay, B.B.; Paranchych, W.; Paetkau, V.H.; Bleackley, R.C.
Science 232, 858-861, 1986

A:Title: Novel serine proteases encoded by two cytotoxic T lymphocyte-specific genes.
 A:Reference number: A94288; MID:86208120
 A:Accession: A94288
 A:Molecule type: mRNA
 A:Residues: 1-247 <LOB>
 A:Cross-references: GB:M12302; NID:9192452; PIDN:AAA37383.1; PID:9309154
 R:Brunet, J.F.; Dosseto, M.; Denizot, F.; Mattei, M.G.; Clark, W.R.; Haeghe, T.M.; Ferrer, P.; Nabholz, M.; Schmitt-Vernhulst, A.M.; Luciani, M.F.; Golstein, P.
 Nature 322, 268-271, 1986
 A:Title: The inducible cytotoxic T-lymphocyte-associated gene transcript CTLA-1 sequence and gene localization to mouse chromosome 14.
 A:Reference number: A93382; MID:86284960
 A:Contents: chromosome mapping
 A:Accession: A93382
 A:Molecule type: mRNA
 A:Residues: 1-247 <BRU>
 A:Cross-references: EMBL:X04072; NID:950586; PIDN:CAA37715.1; PID:950587
 R:Lobe, C.G.; Upton, C.; Dugan, B.; Letellier, M.; Bell, J.; McFadden, G.; Blackley, R.C.
 Biochemistry 27, 6941-6946, 1988
 A:Title: Organization of two genes encoding cytotoxic T lymphocyte-specific serine proteases CCP1 and CCP2.
 A:Reference number: A90536; MID:89062424
 A:Accession: A28952
 A:Molecule type: DNA
 A:Residues: 1-247 <LO2>
 A:Cross-references: GB:M22526; NID:9226432; PIDN:AAB61756.1; PID:9201027
 R:Masson, D.; Tschopp, J.
 Cell 49, 679-685, 1987
 A:Title: A family of serine esterases in lytic granules of cytolytic T lymphocytes
 A:Reference number: A90894; MID:87215932
 A:Accession: B26944
 A:Molecule type: protein
 A:Residues: 21-40 <MAS>
 R:Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maeraki, Y.; Nadeau, J.H.
 Mamm. Genome 5, 349-355, 1994
 A:Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
 A:Reference number: I48934; MID:94319082
 A:Accession: I48937
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 227-247 <RES>
 A:Cross-references: EMBL:U05707; NID:9497035; PIDN:AAB60470.1; PID:9497036
 C:Comment: This enzyme is probably necessary for target cell lysis in cell-mediated immune responses.
 C:Genetics:
 A:Gene: CTLA-1
 A:Map position: 14
 A:Insertions: 19/1; 68/2; 113/3; 200/3
 C:Superfamily: trypsin; serine proteinase; T-cell
 C:Keywords: hydrolase; serine proteinase; T-cell
 F:18/Domain: signal sequence #status predicted <SIG>
 F:19/20/Domain: propeptide #status predicted <APT>
 F:21-24/Product: cytotoxic T-lymphocyte proteinase 1 #status predicted <MPT>
 F:21-24/Domain: cytoxic T-lymphocyte proteinase 1 #status predicted <MPT>
 F:49-65,142-209,173-188/Disulfide bonds: #status predicted
 F:64,108,203/Active site: His, Asp, Ser #status predicted
 PRMSL Length: 247 February 11, 2000 15:51 Type: P Check: 92 ..

1 MKILLILLTL SLASRTKAGE IIGGHEVRPH SRPYMALLSI KDQOEATIG
 51 GFLIREDFVL TAAHCEGSI I NVTLGAHNK EOKETQOIVP MVKCIPIHBDY
 101 NPKTFSDIM LRLKSKAKR TRAVRPLNP RRVNVKPGD VCVYAGWGRM
 151 APMGYSTNL QEVLETFVQKD RECESYFKNR YNKTNQICAG DKPTRASER
 201 GDSGPIVCK KVAAGIVSYG YKDGSPRAF TKVSSFLSWI KTKMSS

!!AA_SEQUENCE 1.0
 P1:OKSE - beta-lactamase (EC 3.5.2.6) precursor - Serratia marcescens
 C:Species: Serratia marcescens
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 18-Jun-1999
 C:Accession: A48176; S11710
 R:Nomura, K.; Yoshida, T.
 FMS Microbiol. Lett. 70, 295-300, 1990
 A:Title: Nucleotide sequence of the Serratia marcescens sr50 chromosome ampC beta-lactamase gene.
 A:Reference number: A48176
 A:Accession: A48176
 A:Molecule type: DNA
 A:Residues: 1-376 <NOM>
 A:Cross-references: EMBL:X52964; NID:947223; PIDN:CAA37137.1; PID:947224
 A:Note: Submitted to the EMBL Data Library May 1990
 A:Note: the authors translated the codon GGA for residue 258 as Asp
 C:Genetics:
 A:Gene: ampC
 C:Superfamily: Escherichia coli beta-lactamase
 C:Keywords: antibiotic resistance; hydrolase
 F:1-11/Domain: signal sequence #status predicted <SIG>
 F:22-376/Product: beta-lactamase #status predicted <MAT>
 F:79/Active site: Ser #status predicted
 OKSE Length: 376 February 11, 2000 15:51 Type: P Check: 107 ..

1 MTKMRCAL IAAIIPFAH AAQOODIDAV IQPIMKRYGV PGMAIIVSD
 51 GKQIIPYGV ASKOTGKPI EQLTEVGS LSKFTATLAV YAOOSKLSF
 101 KDPASHYLPD VRGAFDGV LNLATHTSG LPLFVDDVY NNAOLMAYR
 151 AMQPRHPSGS YRYSNLGIS MGMTAAKSL DQPFQAMQ GMLPALGKSH
 201 TYGVPAQOM ANYAGYSKD DKPVNVNPGP LDAESYGKS NARDIRYLD
 251 ANLQVXYAS VARRPRTS VITSAGFTQ DLWENYPPV VKLSLIEGN
 301 NAGMIMNTP ATATPPQPE LRAGWYNTG STGGFTYAV FIPAKNIVE
 351 MIANKPFND DRYDAVYHI QALEKR
 !!AA_SEQUENCE 1.0
 P1:DS312 - cytidine deaminase (EC 3.5.4.5) - Mycoplasma pirum (strain BER) (SGC3)
 C:Species: Mycoplasma pirum
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: D53312
 R:Tham, T.N.; Ferris, S.; Kovacic, R.; Montagnier, L.; Blanchard, A.
 J. Bacteriol. 175, 5281-5285, 1993
 A:Title: Identification of Mycoplasma pirum genes involved in the salvage pathways for nucleosides.
 A:Reference number: A53312; MID:93352438
 A:Accession: D53312
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-133 <THA>
 A:Cross-references: GB:L13289; NID:9401781; PIDN:AAA25433.1; PID:9401785
 C:Genetics:
 A:Gene: code: SGC3
 C:Superfamily: cytidine deaminase
 C:Keywords: hydrolase
 D53312 Length: 133 February 11, 2000 15:51 Type: P Check: 6718 ..

1 MKKDIYQK INELISNAVY PYSNFRVSC LITDGNFAG VNIENSAYS
 51 TICAERSAVS SMITSGEQI FRVYILTDTI VKDIGPCGV CROVLSERAK
 101 PETPIITYNL KGEFFYTLLE QLLPFAFNKD ALX
 !!AA_SEQUENCE 1.0

P1:JE0022 - cytidine/deoxycytidine deaminase (EC 3.5.4.-) cdd - Bacillus subtilis
N:Alternate names: cytidine 12'-deoxycytidine aminohydrolase; P43 protein
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JE0022; I39955; F69597; S04425
R:Song, B.H.; Neubard, J.
Mol.Gen. Genet. 216: 462-468, 1989
A:Title: Chromosomal location, cloning and nucleotide sequence of the Bacillus subtilis cdd gene encoding cytidine/deoxycytidine deaminase.
A:Reference number: JE0022; M01D:8931367
A:Accession: JE0022
A:Molecule type: DNA
A:Residues: 1-136 <SON>
A:Cross-references: GB:018532; MID:9606743; PIDN:AAB59993.1; PID:9606744
R:Wang, P.
J. Biol. Chem. 259, 8619-8625, 1984
A:Title: Overlapping promoters transcribed by Bacillus subtilis sigma-55 and sigma-37 RNA polymerase holoenzymes during growth and stationary phases.
A:Reference number: I39955; M01D:84239852
A:Accession: I39955
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-10 <RES>
A:Cross-references: GB:R02174; MID:9143349; PIDN:AAB05347.1; PID:9551724
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bartoe, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, C.V.; Caldwel, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Delizet, F.; Devine, K.M.; Dusterheft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Farber, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D., Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Guseppl, G.; Guy, B.O.; Hage, K.; Halech, U.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hultlo, M.F.; Itaya, M.; Jones, L.; Jorits, B.; Karamata, D.; Kashimura, Y.; Kieffer-Bleichard, M.; Klein, C.; Kodayashi, Y.; Koether, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Minano, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogizawa, A.; Ouega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porrolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.
A:Authors: Schleich, S.; Schreuter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, P.; Vassalotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Wetzinger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Darchin, A.
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; M01D:98044033
A:Accession: F69597
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <KUN>
A:Cross-references: GB:299116; GB:299117; GB:AL009126; MID:92634966; PIDN:AB144472.1; PID:el183760; PID:92634976; MID:92634973; PID:el185796; PID:92634962
A:Experimental source: strain 168
C:Genetics:
A:Gene: cdd
A:Map position: 225 (degrees)
C:Superfamily: cytidine deaminase
C:Keywords: hydrolase
JE0022 Length: 136 February 11, 2000 15:51 Type: P Check: 9048 ..

11AA_SEQUENCE 1.0
P1:F69500 - riboflavin-specific deaminase (ribd) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1998 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: F69500
R:Kliam, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenry, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Goehyne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Ullrich, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujita, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
A:Reference number: A69250; M01D:98049343
A:Accession: F69500
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-219 <KLE>
A:Cross-references: GB:AE000964; GB:AE000782; MID:92669287; PIDN:AAB89247.1; PID:92648530; TIGR:AF2007
C:Superfamily: riboflavin-specific deaminase
F69500 Length: 219 February 11, 2000 15:51 Type: P Check: 4460 ..

1 MRPYFVAVVA ASDGKTSDE SKRLRISCE EDLKVDRLR AESDAIMVGI
51 GTVLADPRL TVKSALIREK ROKDGKEPNP LRVVDSRCR VPIIARLND
101 EARLVAISR IADBEKREV KVAEVAVFG EERYELALL EFLHRGVR
151 LMVEGGGLLI SSLISONLVD EIRIYGPFI IGRHDSPTVC DGBSLKCR
201 IEKIERIGEG FAVTARENR

11AA_SEQUENCE 1.0
P1:18612 - H+-transporting ATP synthase (EC 3.6.1.34) lipid-binding protein P3 precursor, mitochondrial - human
N:Alternate names: H+-transporting ATP synthase chain 9.3; H+-transporting ATP synthase chain c form P3
C:Species: Homo sapiens (hmn)
C:Date: 26-Aug-1999 #sequence_revision 26-Aug-1999 #text_change 26-Aug-1999
C:Accession: 18612
R:Ryan, M.L.; Ierner, T.J.; Haines, J.L.; Gusella, J.F.
Genomics 24, 375-377, 1994
A:Title: Sequence analysis and mapping of a novel human mitochondrial ATP synthase subunit 9 cDNA (ATP53)
A:Reference number: A55677; M01D:95213032
A:Accession: 18612
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-142 <YAN>
A:Cross-references: EMBL:U09813; MID:91008454; PIDN:AAA78807.1; PID:9511450
C:Genetics:
A:Gene: ATP5G3
A:Cross-references: GDB:375306
A:Map position: 2pter-2qter
A:Genome: nuclear
C:Superfamily: H+-transporting ATP synthase lipid-binding protein
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex; mitochondrion; oxidative phosphorylation; transmembrane protein

F:1-67/Domain: transit peptide (mitochondrion) #status predicted <TRP>
 F:68-142/Product: H+-transporting ATP synthase lipid-binding protein #status
 predicted <MAT>
 F:75-103/Domain: transmembrane #status predicted <TM1>
 F:113-142/Domain: transmembrane #status predicted <TM2>
 F:125/Active site: Glu #status predicted

I38612 Length: 142 February 11, 2000 15:51 Type: P Check: 2178 ..

1 MFACACIACF PSLIRAGSRV AYRISASV SRPEASRGE GSTYENQAN
 51 GVSOLIERE OTSAISRDI TAAFTGAG AIVGAGSQA GIGVFQSLI
 101 IGVARNPSLK QOLFSTAILG FALSEANGF CLMAVFLILF AM

11AA_SEQUENCE 1.0
 P1:RKTOS2 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain 2
 precursor - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 18-Jun-1999
 C:Accession: S02363; B24885; B29037; T07157
 R:Sugita, M.; Manzara, T.; Pichersky, E.; Cashmore, A.; Grussem, W.
 Mol. Gen. Genet. 209, 247-256, 1987
 A:Title: Genomic organization, sequence analysis and expression of all five
 genes encoding the small subunit of ribulose-1,5-bisphosphate
 carboxylase/oxygenase from tomato.
 A:Reference number: S01107; M01D:86038372
 A:Accession: S02363
 A:Molecule type: DNA
 A:Residues: 1-180 <SUG>
 A:Cross-references: EMBL:X05983
 A>Note: The authors translated the codon TCA for residue 32 as Thr and ACT for
 residue 34 as Ser
 R:Pichersky, E.; Bernatzky, R.; Tanksley, S.D.; Cashmore, A.R.
 Proc. Natl. Acad. Sci. U.S.A. 83, 3880-3884, 1986
 A:Title: Evidence for selection as a mechanism in the concerted evolution of
 Lycopersicon esculentum (tomato) genes encoding the small subunit of
 ribulose-1,5-bisphosphate carboxylase/oxygenase.
 A:Reference number: A24885; M01D:86233336
 A:Accession: B24885
 A:Molecule type: mRNA
 A:Residues: 1-180 <PIC>
 A:Cross-references: EMBL:M13543; NID:g170497; PIDN:AAA34189.1; PID:g170498
 A>Note: the authors translated the codon CTC for residue 5 as Ala
 R:McNigh, T.D.; Alexander, D.C.; Babcock, M.S.; Simpson, R.B.
 Gene 48, 23-32, 1986
 A:Title: Nucleotide sequence and molecular evolution of two tomato genes
 encoding the small subunit of ribulose-1,5-bisphosphate carboxylase.
 A:Reference number: A29037; M01D:87163513
 A:Accession: B29037
 A:Molecule type: mRNA
 A:Residues: 1-86, 'V', 88-180 <MCK>
 A:Cross-references: EMBL:M15236; NID:g170503; PIDN:AAA34192.1; PID:g170504
 R:Manzara, T.
 submitted to the EMBL Data Library, August 1989
 A:Reference number: 215964
 A:Accession: T07157
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-180 <MAN>
 A:Cross-references: EMBL:X05983; NID:g19331; PIDN:CAA29401.2; PID:e1389536
 A:Experimental source: cultivar VNT LA1221 cherry line
 C:Genetics:
 A:Gene: rbcS-2
 A:Map position: 3
 A:Introns: 59/3; 104/3; 122/2
 C:Superfamily: ribulose-bisphosphate carboxylase small chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;
 carboxy-lyase; chloroplast; monooxygenase; photorespiration
 F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:58-180/Product: ribulose-bisphosphate carboxylase small chain 2 #status
 predicted <MAT>

RKTOS2 Length: 180 February 11, 2000 15:51 Type: P Check: 4535 ..

1 MASSIVSSAA VATRSNVTQA SMVAPFTGLK SSATFPVTK ONLDTISAS
 51 NGRVSCMOV WPIINKKYE TLYSLPDLSD EQLLSEIEY LKNGVPCLE
 101 FETHEGFYR ENKSPGYD GRVWTMKLP MFCCTDATOV LAEVOEAKKA
 151 YPOAWVRITG FDNVRQVOCI SFIAKPEGY

11AA_SEQUENCE 1.0
 P1:RKTOS1 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain 1
 precursor - tomato
 C:Species: Lycopersicon esculentum (tomato)
 C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 18-Jun-1999
 C:Accession: S02364; A24885; A29037; S21583
 R:Sugita, M.; Manzara, T.; Pichersky, E.; Cashmore, A.; Grussem, W.
 Mol. Gen. Genet. 209, 247-256, 1987
 A:Title: Genomic organization, sequence analysis and expression of all five
 genes encoding the small subunit of ribulose-1,5-bisphosphate
 carboxylase/oxygenase from tomato.
 A:Reference number: S01107; M01D:86038372
 A:Accession: S02364
 A:Molecule type: DNA
 A:Residues: 1-181 <SUG>
 A:Cross-references: EMBL:X05982; NID:g19326; PIDN:CAA29400.1; PID:g295814
 A>Note: the authors translated the codon CTT for residue 49 as Ile
 R:Pichersky, E.; Bernatzky, R.; Tanksley, S.D.; Cashmore, A.R.
 Proc. Natl. Acad. Sci. U.S.A. 83, 3880-3884, 1986
 A:Title: Evidence for selection as a mechanism in the concerted evolution of
 Lycopersicon esculentum (tomato) genes encoding the small subunit of
 ribulose-1,5-bisphosphate carboxylase/oxygenase.
 A:Reference number: A24885; M01D:86233336
 A:Accession: A24885
 A:Molecule type: DNA
 A:Residues: 1-181 <PIC>
 A:Cross-references: EMBL:M13542; NID:g170495; PIDN:AAA34188.1; PID:g170496
 R:McNigh, T.D.; Alexander, D.C.; Babcock, M.S.; Simpson, R.B.
 Gene 48, 23-32, 1986
 A:Title: Nucleotide sequence and molecular evolution of two tomato genes
 encoding the small subunit of ribulose-1,5-bisphosphate carboxylase.
 A:Reference number: A29037; M01D:87163513
 A:Accession: A29037
 A:Molecule type: mRNA
 A:Residues: 1-181 <MCK>
 A:Cross-references: EMBL:M15235
 R:Manzara, T.; Carrasco, P.; Grussem, W.
 submitted to the EMBL Data Library, April 1992
 A:Description: Developmental and organ-specific changes in DNA-protein
 interactions in the tomato rbcS1, rbcS2 and rbcS3A promoter regions.
 A:Reference number: S21582
 A:Accession: S21583
 A:Molecule type: DNA
 A:Residues: 1-20 <MAN>
 A:Cross-references: EMBL:X66068; NID:g22621; PIDN:CAA46868.1; PID:g22622
 C:Genetics:
 A:Gene: rbcS-1
 A:Map position: 2
 A:Introns: 60/3; 105/3
 C:Superfamily: ribulose-bisphosphate carboxylase small chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase;
 carboxy-lyase; chloroplast; monooxygenase; photorespiration
 F:58/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:59-181/Product: ribulose-bisphosphate carboxylase small chain 1 #status
 predicted <MAT>

RKTOS1 Length: 181 February 11, 2000 15:51 Type: P Check: 5240 ..

1 MASSIVSSAA AATRSNVAQA SWVAPFTGLK SAASEPVTK NNVVDITSLA
 51 SNGRVRCMO VWPINKKRY ETLSTYPLDS DEQLSEIEY LKNGVPCLE
 101 EFERGERFYR RENNSSPGY DGRVWTMKL PMFCCTATOV VIAEVOEAKK

151 AYPQAWIRII GEDNVRQVOC ISFIAYKPEG F

!!AA_SEQUENCE 1.0
 P1:RKPOS8 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) precursor small chain rbcS-1 - potato
 C:Species: Solanum tuberosum (potato)
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
 C:Accession: A31083
 R:Moeller, F.P.; Fritsch, C.C.; Willmitzer, L.; Schell, J.; Schreier, P.H.
 Proc. Natl. Acad. Sci. U.S.A. 85, 846-850, 1988
 A:Title: rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution.
 A:Reference number: A31083; MUID:86124937
 A:Accession: A31083
 A:Molecule type: mRNA
 A:Residues: 1-181 <MW>
 A:Cross-References: GB:J03613; NID:g169556; PIDN:AAA3838.1; PID:g169557
 A:Experimental source: strain H1201/7
 C:Comment: Ribulose-bisphosphate carboxylase, a major component of leaf protein, is also a monooxygenase; it catalyzes the carboxylation of D-ribulose 1,5-bisphosphate (the primary event in photosynthetic carbon dioxide fixation) as well as the oxidative fragmentation of the pentose substrate in the photorespiration process. These reactions occur simultaneously and in competition at the same active site.
 C:Comment: Each active molecule contains eight large chains, synthesized on the chloroplast ribosomes and containing the active site, and eight small chains, the precursors of which are synthesized on cytoplasmic ribosomes and converted to mature small chains during or immediately after transport into the chloroplast.
 C:Comment: This protein is coded by one member of a small multigene family.
 C:Genetics:
 A:Gene: rbcS-c
 A:Superfamily: ribulose-bisphosphate carboxylase small chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
 F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:58-181/Product: ribulose-bisphosphate carboxylase small chain rbcS-c #status predicted <MAT>

RRPOS8 Length: 181 February 11, 2000 15:51 Type: P Check: 4699 ..

1 MASSIVSSAA VATRSNVAQA SMVAPFTGLK SAASFPVTRK NNNVDITSLA

51 SNGGRVRCMQ WMPPIIMKKY ETLSTLPDLT DEQLKEVEY LKNGWVPC

101 EETEHGFEVY REHNSSPGY DGRWTMVKL PMFGCTDGTQ VLAEOEAKN

151 AYPQAWIRII GEDNVRQVOC ISFIAYKPEG Y

!!AA_SEQUENCE 1.0
 P1:RKPOS1 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) precursor small chain rbcS-1 - potato
 C:Species: Solanum tuberosum (potato)
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
 C:Accession: B31083; S31497
 R:Moeller, F.P.; Fritsch, C.C.; Willmitzer, L.; Schell, J.; Schreier, P.H.
 Proc. Natl. Acad. Sci. U.S.A. 85, 846-850, 1988
 A:Title: rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution.
 A:Reference number: A31083; MUID:86124937
 A:Accession: B31083
 A:Molecule type: DNA
 A:Residues: 1-181 <MW>
 R:Fritsch, C.C.; Moeller, F.P.; Schenkemeyer, V.; Hergel, T.; Schreier, P.H.
 submitted to the EMBL Data Library, December 1992
 A:Reference number: S31494
 A:Accession: S31497
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-181 <FRI>
 A:Cross-References: EMBL:X69759; NID:g21562; PIDN:CAA9413.1; PID:g21563
 C:Genetics:

A:Gene: rbcS-1
 A:Introns: 59/3; 102/3
 C:Superfamily: ribulose-bisphosphate carboxylase small chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
 F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:58-181/Product: ribulose-bisphosphate carboxylase small chain rbcS-1 #status predicted <MAT>

RRPOS1 Length: 181 February 11, 2000 15:51 Type: P Check: 4663 ..

1 MASSIVSSAA VATRNVTAQA GSMVAPFTGLK KSATFPVSR KNNLDITSLA

51 SNGGRVRCMQ WMPPIIMKKY ETLSTLPDLT DEQLKEVEY LKNGWVPC

101 EETEHGFEVY REHNSSPGY DGRWTMVKL PMFGCTDGTQ VLAEOEACK

151 SYPQAWIRII GEDNVRQVOC ISFIAYKPEG Y

!!AA_SEQUENCE 1.0
 P1:RRPOS8 - ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain (ssu8) precursor - garden petunia
 C:Species: Petunia x hybrida (garden petunia)
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1999
 C:Accession: A24917
 R:Turner, N.E.; Clark, W.G.; Tabor, G.J.; Hironaka, C.M.; Fraley, R.T.; Shah, D.N.
 Nucleic Acids Res. 14, 3325-3342, 1986
 A:Title: The genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase are expressed differentially in petunia leaves.
 A:Reference number: A93619; MUID:86205237
 A:Accession: A24917
 A:Molecule type: DNA
 A:Residues: 1-180 <TMP>
 A:Cross-References: GB:X03820; NID:g20490; PIDN:CAA2444.1; PID:g20491
 C:Genetics:
 A:Gene: rbcS
 A:Introns: 59/3; 104/3; 122/3
 C:Superfamily: ribulose-bisphosphate carboxylase small chain
 C:Keywords: Calvin cycle; carbon dioxide fixation; carbon-carbon lyase; carbon-oxygen lyase; carboxy-lyase; chloroplast; monooxygenase; photorespiration
 F:1-57/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F:58-180/Product: ribulose-bisphosphate carboxylase small chain (ssu8) #status predicted <MAT>

RRPOS8 Length: 180 February 11, 2000 15:51 Type: P Check: 2967 ..

1 MASSIVSSAA VATRTNVAQA SMVAPFNGLK SAASFPVSRK ONNLDITSLA

51 NGRVRCMQW WPPYGRKKYE TLTSLPDLT DEQLKEVEY LKNGWVPC

101 EETEHGFEVY EYHASPRYD GRWTMVKLP MFECTDGTQV LGELDEAKA

151 YPNAMIRIIG FDNVRQVOCI SFIAKPPGF

!!AA_SEQUENCE 1.0
 P1:F69398 - dihydroxyxynaphthohol acid synthase (menB) homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: F69398
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gaitan, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervilavage, A.R.; Graham, D.E.; Kipridis, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.F.; Dougherty, B.A.; McHenry, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reisch, C.I.; McNeill, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic

sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 A:Reference number: A69250; MUID:98049343
 C:Species: F09398
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-277 <KLE>
 A:Cross-references: GB:AE001022; GB:AE000782; NID:g2689345; PIDN:AA90054.1;
 PID:g2649394; TIGR:AF1191
 C:Superfamily: naphthate synthase; enoyl-CoA hydratase homology
 F:29-192/Domain: enoyl-CoA hydratase homology <ECH>
 F69398 Length: 277 February 11, 2000 15:51 Type: P Check: 4436 ..

1 MGFLFODII YEKEGRVAKI TNRPEKINA CTPTVVEIS KAFIDAMTDR
 51 KIGVVVETGA GDKAFVGGD OSIRLDGYS YSSEELGII ALPLEVGMQ
 101 IYFLLRHIP KPIAVRNGY AVGGHVMQV NCDLSIAER AKFGQAPRV
 151 GSFDFGFGT ELNRVGMKR AKEIFLCEL YTAELALMG LNAVVPHEK
 201 LDEVEKMS ELLEKSPAL KMLKYAFLAD TEGLAGITEL GVGGLSLYGG
 251 TRESLEGRNA FMEKRKDFW KEVEGGD

11AA-SEQUENCE 1.0
 PI:F71428 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: Columbia
 C>Date: 10-Sep-1998 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: F71428
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergamp, R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel, K.; Murphy, G.; Piffanelli, P.; Medler, H.; Medler, E.; Wandt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Gieslen, S.; Villarroel, R.; de Clerck, R.; Van Montagu, M.; Lecharny, A.; Adoozy, S.; Gy, I.; Kreis, N.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pildomenech, P.; Douka, A.; Vouklatou, E.; Millon, D.; Hatzopoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reckman, S.; Ansoorge, W.; Cooke, R.; Berger, C.; Delseny, M.; Voelt, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.; Schueller, C.; Chaiwatizis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana.
 A:Reference number: A71400; MUID:98121113
 A:Accession: F71428
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-244 <BNV>
 A:Cross-references: GB:297340; NID:g2244950; PIDN:CAB10400.1; PID:e327497;
 PID:8224979
 C:Genetics:
 A:Map position: 4COP9-4G3845
 C:Superfamily: naphthate synthase; enoyl-CoA hydratase homology
 F:31-180/Domain: enoyl-CoA hydratase homology <ECH>
 F71428 Length: 244 February 11, 2000 15:51 Type: P Check: 9250 ..

1 MDOTVSENLI QVKEESGIA VITINPKSL NSLTRAMVD LARAFKDNDS
 51 DESVQVVIPT GSGRFSFGV DLTAAASVR GDVKNPDPD VYOMERLARP
 101 IIGAINGFAT TAGFELALAC DILVASRGAK FMDTHARFGI FPSWGLSOKL
 151 SRIIGANKAR EVSLTSMPLT ADVAGKLGKV NHVVEGEGAL KKAREIAEAI
 201 IKNEQGVLR IKSVINDLK LDGHAULTLE KVKHSAIENL LINDL

11AA-SEQUENCE 1.0
 PI:A48681 - placental thrombin inhibitor - human

N:Alternate names: cytoplasmic antiprotease; intracellular serine proteinase inhibitor, 38k
 C:Species: Homo sapiens (man)
 C>Date: 07-Apr-1994 #sequence_revision 07-Jul-1995 #text_change 18-Jun-1999
 C:Accession: A48681; A54352; A46672; B46672; C46672; S35750
 R:Conklin, P.; Sun, J.; Cerretti, L.; Salem, H.H.; Bird, P.
 Proc. Natl. Acad. Sci. U.S.A. 90, 9417-9421, 1993
 A:Title: Cloning and molecular characterization of a human intracellular serine proteinase inhibitor.
 A:Reference number: A48681; MUID:94022386
 A:Accession: A48681
 A:Molecule type: mRNA
 A:Residues: 1-376 <COO>
 A:Cross-references: GB:222658; NID:g297411; PIDN:CA80373.1; PID:g297412
 A:Experimental source: Placenta
 A:Note: authors translated the codon CAA for residue 198 as Gly
 R:Morgenstern, K.A.; Sprecher, C.; Holth, L.; Foster, D.; Grant, F.J.; Ching, A.; Kiesel, W.
 Biochemistry 33, 3432-3441, 1994
 A:Title: Complementary DNA cloning and kinetic characterization of a novel intracellular serine proteinase inhibitor: mechanism of action with trypsin and factor Xa as model proteinases.
 A:Reference number: A54352; MUID:94183847
 A:Accession: A54352
 A:Molecule type: mRNA
 A:Residues: 1-174; 'E', 176-361; 'S', 363-376 <MOR>
 A:Cross-references: GB:569272; NID:g546087; PIDN:AA830320.1; PID:g546088
 A:Experimental source: Placenta
 A:Note: sequence extracted from NCBI backbone (NCBI:145231, NCBI:145232)
 R:Conklin, P.B.; Tetaz, T.; Salem, H.H.
 J Biol Chem 268, 9541-9547, 1993
 A:Title: Identification and purification of a novel serine proteinase inhibitor.
 A:Reference number: A46672; MUID:93252826
 A:Accession: A46672
 A:Molecule type: Protein
 A:Residues: 47-60; 63-81; 91-98 <CO2>
 A:Experimental source: Placenta, leukemic cell line K562
 A:Note: sequence modified after extraction from NCBI backbone
 C:Genetics:
 A:Gene: GDB:PI6
 A:Cross-references: GDB:252025; OMIM:173321
 A:Map position: 6p25-6p24.3
 C:Superfamily: antithrombin III
 C:Keywords: blocked amino end; cytosol; serine proteinase inhibitor
 F:341/inhibitory site: Arg (thrombin) #status predicted
 A48681 Length: 376 February 11, 2000 15:51 Type: P Check: 300 ..

1 MDVLAENGT FALNLKTLG KDNKNVFFS PMSKSCALAM VYMGAKGNTA
 51 AQMAQLISFN KSGGGGDHQQ GFOSLITEVN KTGYOYLRY ANRLFGKSC
 101 DFLSFRDSC QKFOAEEMEE LDFISAVES RKHINTVVAE KTEGKIAELL
 151 SPGSVPLFR LVLVNAVYFR GNMDDGOFDE NTEERLEKVS KNEKPYQM
 201 FKSTFKKTY IGEIFQDILV LPYVGEKLMN IIMPDETDT LRYVELELY
 251 EKVEVTRFD MMDDEVEVS LPRFKLEESY DMESVLNLG MDAFLGLGA
 301 DSGMSQDPL SLKVKHKSF VEYNEGCTEA AAATRAIMM RCARFVPRC
 351 ADHPFLFTIO HRKTNGLILFC GRFSSP

11AA-SEQUENCE 1.0
 PI:W17WA - alpha-amylase inhibitor CII - wheat
 C:Species: Triticum aestivum (common wheat)
 C>Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 04-Sep-1998
 C:Accession: A01322
 R:Kashlan, N.; Richardson, M.
 Phytochemistry 20, 1781-1784, 1981
 A:Title: The complete amino acid sequence of a major wheat protein inhibitor of alpha-amylase.

A:Reference number: A01322
A:Accession: A01322
A:Molecule type: protein
A:Residues: 1-123 <KAS>
A:Note: half of the molecules contained 65-Ser. Very small amounts of 67-Ala, 98-Gly, 99-Pro, and 118-Val were also found.
C:Comment: This protein is one of a number of related alpha-amylase inhibitors found in the endosperm of wheat seeds.
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha amylase inhibitor
F:1-52,21-42,29-82,43-98,56-113/Disulfide bonds: #status predicted
MIWTA Length: 123 February 11, 2000 15:51 Type: P Check: 4689 ..

1 SGPWMCNPA TGYKVSALTG CRAMKLCY GSOVPAVLK DCCOQLADIN
51 NEMCRGDIS SMLRAYOEL GVEGKEVLE GCRKEVMKLT AASYPEVCXY
101 PIPNPSGDRA GVCYGDWCAY PDV

!!AA_SEQUENCE 1.0
PI:MIWTA5 - alpha-amylase inhibitor 0.53 - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 13-Jun-1983 #sequence_revision 28-Feb-1986 #text_change 04-Sep-1998
C:Accession: A90649; A90656; A01323
R:Maeda, K.; Hase, T.; Matsubara, H.
Biochim. Biophys. Acta 743, 52-57, 1983
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat kernel
A:Reference number: A90649; MUID:83127436
A:Accession: A90649
A:Molecule type: protein
A:Residues: 1-119, YPPKA, <MA>
A:Note: this sequence has been revised in reference A90656
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
Biochim. Biophys. Acta 828, 213-221, 1985
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat kernel (0.19-inhibitor).
A:Reference number: A90656; MUID:85175148
A:Accession: A90656
A:Molecule type: protein
A:Residues: 1-124 <MA>
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
J. Biochem. 94, 865-870, 1983
A:Title: Disulfide bridges in an alpha-amylase inhibitor from wheat kernel.
A:Reference number: A91976; MUID:84061717
A:Accession: A91976
A:Contents: annotation: disulfide bonds
C:Comment: This protein is a dimer of identical chains.
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha-amylase inhibitor
F:20-41,42-99,54-115/Disulfide bonds: #status predicted
F:28-83/Disulfide bonds: #status experimental
MIWTA5 Length: 124 February 11, 2000 15:51 Type: P Check: 4073 ..

1 SGPWMCYPCQ AFQVPALPGC RPLIKLCNG SOVPEAVLRD CCOQLADISE
51 WPRGALYSM LDSMKREHG SEGOAGTGAF PSGRREVVKL TAASITAVCR
101 LPIVDASGD GAYVCKDVAA YPDA

!!AA_SEQUENCE 1.0
PI:MIWTA1 - alpha-amylase inhibitor 0.19 - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 28-Feb-1986 #sequence_revision 28-Feb-1986 #text_change 30-Sep-1998
C:Accession: A01324
R:Maeda, K.; Wakabayashi, S.; Matsubara, H.
Biochim. Biophys. Acta 828, 213-221, 1985
A:Title: Complete amino acid sequence of an alpha-amylase inhibitor in wheat kernel (0.19-inhibitor).
A:Reference number: A90656; MUID:85175148
A:Accession: A01324
A:Molecule type: protein

A:Residues: 1-124 <MA>
R:Okuda, M.; Satoh, T.; Sakurai, N.; Shibuya, K.; Kaji, H.; Samejima, T.
J. Biochem. 122, 918-926, 1997
A:Title: Overexpression in Escherichia coli of chemically synthesized gene for active 0.19 alpha-amylase inhibitor from wheat kernel.
A:Reference number: JC5705; MUID:98104043
A:Accession: JC5705
A:Contents: annotation
C:Comment: For the sequence from a synthetic DNA used for artificial protein production, see PIR:JC5705.
C:Complex: homodimer
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: alpha amylase inhibitor; homodimer
F:8-52,20-41,28-83,42-99,54-115/Disulfide bonds: #status predicted
MIWTA1 Length: 124 February 11, 2000 15:51 Type: P Check: 3320 ..

1 SGPWMCYPCQ AFQVPALPGC RPLIKLCNG SOVPEAVLRD CCOQLADISE
51 WCRGALYSM LDSMKREHGA QEGOAGTGAF PSGRREVVKL TAASITAVCR
101 LPIVDASGD GAYVCKDVAA YPDA

!!AA_SEQUENCE 1.0
PI:TVHUF1 - transforming protein fos - human
C:Species: Homo sapiens (man)
C:Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 24-Sep-1999
C:Accession: A01342; I58359; S09375
R:van Straaten, F.; Muller, R.; Curran, T.; Van Beveren, C.; Verma, I.M.
Proc. Natl. Acad. Sci. U.S.A. 80, 3183-3187, 1983
A:Title: Complete nucleotide sequence of a human c-onc gene: deduced amino acid sequence of the human c-fos protein
A:Reference number: A01342; MUID:83221560
A:Accession: A01342
A:Molecule type: DNA
A:Residues: 1-360 <VAN>
A:Cross-references: GB:V01512; NID:929903; PIDN:CAA24756.1; PID:929904
R:Nakabeppu, Y.; Nathans, D.
EMBO J. 8, 3833-3841, 1989
A:Title: The basic region of Fos mediates specific DNA binding.
A:Reference number: S09374; MUID:9005986
A:Accession: S09374
A:Contents: annotation
A:Note: engineered chimeric forms including residues 121-207
R:Roux, P.; Verrier, B.; Klein, B.; Niccolino, M.; Marty, L.; Alexandre, C.; Piechaczyk, M.
Oncogene 6, 2155-2160, 1991
A:Title: Retrovirus-mediated gene transfer of a human c-fos cDNA into mouse bone marrow stromal cells.
A:Reference number: I58359; MUID:92050815
A:Accession: I58359
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-6 <RSS>
A:Cross-references: GB:S65138; NID:9238776; PIDN:AAB20306.1; PID:9238778
C:genetics:
A:Gene: GDB:PO5
A:Cross-references: GDB:119917; OMIM:164810
A:Map position: 14q24.3-14q24.3
A:Introns: 47/3; 131/3; 167/3
C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; heterodimer; leucine zipper
F:132-172/Domain: fos/jun DNA-binding domain homology <FUD>
F:165-193/Region: leucine zipper motif
TVHUF1 Length: 360 February 11, 2000 15:51 Type: P Check: 4677 ..

1 MWFSGFNADY EASSRCSA SPAGDSLSTY HSPADSFISM GSPVNAQDFC
51 TDLAVSANF IPTVTAISTS PDLQWLQPA LVSSVAPSGT RAHPFEGVPA
101 PSNGATSRAG VYKTMIGRA QSIGRGKYE QLSPEEEKR RIRERKXMA
151 AAKCRNRRE LTRDLQAEIT QLEDEKSAQ TEIANLLEK EKLEFIILAH

```

201 RPAKIPDDL GFPEKSVAS LDLTGLPEV ATPSEBET LPLNDPEPK
251 PSVEPVKIS SMELKTEPD DFLPASP RP SGSETPASVP DMDLSGSTYA
301 ADWEPLHSGS LGMGPATL EPLCTPVTC TPSCATVSS FVFTYHEADS
351 PPSCAAHK GSSSEPPSSD SLSPSLIAL

!!AA_SEQUENCE 1.0
P1:TVFV4 - transforming protein fos - chicken
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 18-Jun-1999
C:Accession: A28368; A34073
R:Moelders, H.; Januwein, T.; Adamkiewicz, J.; Mueller, R.
Oncogene 1, 377-385, 1987
A:Title: Isolation and structural analysis of a biologically active chicken
c-fos cDNA: identification of evolutionarily conserved domains in fos protein.
A:Reference number: A28368; MUID:88262231
A:Accession: A28368
A:Molecule type: mRNA
A:Residues: 1-367 <MOB>
A:Cross-references: GB:M18043; NID:9211469; PIDN:AA48670.1; PID:9211470
K:Fujimura, K.T.; Ashida, K.; Nishina, H.; Iba, H.; Miyajima, N.; Nishizawa,
M.; Kawai, S. 4012-4018, 1987
J:Virol. 61, 4012-4018, 1987
A:Title: The chicken c-fos gene: cloning and nucleotide sequence analysis.
A:Reference number: A34073; MUID:88062957
A:Accession: A34073
A:Molecule type: DNA
A:Residues: 1-367 <FUD>
A:Cross-references: GB:M18043; NID:9211467; PIDN:AA476823.1; PID:9211468
A:Note: the authors translated the codon AAC for residue 90 as Asp
C:Genetics:
A:Gene: fos
A:introns: 47/3; 130/3; 166/3
C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; leucine zipper; transforming protein
F:31-171/Domain: fos/jun DNA-binding domain homology <FUD>
F:164-192/Region: leucine zipper motif

TVCHFS Length: 367 February 11, 2000 15:51 Type: P Check: 7007

1 MMVGFAGEX EAPSRCSA SPAGDILTY PSPADSESSM GSPVNSQDFC
51 TDLAASANF VPTVAISTS PDLQWLVOPT LISSVAFSON RGHAYGVAPAP
101 APPAAISRPV VLKAPGKQVQ SIGRKGKVGQ LSPDEEKRR IRREKMKMA
151 AKCRNRREL TDTLQAEVDQ LEEKSALQA ELANLKEKE KLEFILAHR
201 PACKMEELR FSEELAAATA LDGASPPAA AEAFALPLM TEAPPAVPPK
251 EPGSGLELK AEPDELLFS AGPREASRV PDMULPQSS FYASDWEPLG
301 AGSGGELEPL CTPTVTCPC PSTYTSTFVF TYPEDAPPS CAAAHKRGSS
351 SNEPSSDSLS SPTLIAL

!!AA_SEQUENCE 1.0
P1:TVFV4 - transforming protein fos - avian retrovirus NK24
C:Species: avian retrovirus NK24
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 26-May-1995
C:Accession: B34071
R:Nishizawa, M.; Goto, N.; Kawai, S.
J:Virol. 61, 3733-3740, 1987
A:Title: An avian transforming retrovirus isolated from a nephroblastoma that
carries the fos gene as the oncogene.
A:Reference number: A34071; MUID:88062920
A:Accession: B34071
A:Molecule type: DNA
A:Residues: 1-322 <NIS>
C:Genetics:
A:Gene: fos

C:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
C:Keywords: DNA binding; leucine zipper; transforming protein
F:86-126/Domain: fos/jun DNA-binding domain homology <FUD>
F:119-147/Region: leucine zipper motif

TVFV4 Length: 322 February 11, 2000 15:51 Type: P Check: 9324

1 SODECTDLAV SSANFPTVT AISTSPDLQW LVQFTLLSSV APSQNRGHPY
51 GVPAPAPPA YSRPAVILKAP GGRGOSIGR GKVEQLSPEE EKKRIRRR
101 NKMAAKCN RRRLDTLQ AETDLEBK SALQAEIANT LKEKELEPI
151 LAHRPACRM PEELFSEEL AATLADIGA PSPAAEEAF ALPLTEADP
201 AVPPKEPSGS GLELKAEFD ELLFSAGPRE ASRSVPMDL PGASSFYASD
251 WEPLGASGG ELEPLCTPV TCTPCPSTYT STVFITYEA DAFPSCAAH
301 RKGSSNEPS SDLSPTLL AL

!!AA_SEQUENCE 1.0
P1:S00755 - pleckstrin - human
N:Alternate names: p47; platelet/leukocyte C kinase substrate (pleckstrin)
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-Jul-1999
C:Accession: S00755; A45762
R:Tyers, M.; Haslam, R.A.; Stewart, M.I.; Varrichio, A.M.; Short, R.G.L.;
Haslam, R.J.; Harley, C.B.
Nature 333, 470-473, 1988
A:Title: Molecular cloning and expression of the major protein kinase C
substrate of platelets
A:Reference number: S00755; MUID:88232910
A:Accession: S00755
A:Molecule type: mRNA
A:Residues: 1-350 <TYE>
A:Cross-references: EMBL:X07743; NID:935517; PIDN:CAA30564.1; PID:935518
R:Tyers, M.; Haslam, R.J.; Rachubinski, R.A.; Harley, C.B.
J:Cell. Biochem. 40, 133-143, 1989
A:Title: Molecular analysis of pleckstrin: the major protein kinase C substrate
of platelets.
A:Reference number: A45762; MUID:89359547
A:Accession: A45762
A:Molecule type: mRNA
A:Residues: 1-350 <TYE>
A:Cross-references: GB:X07743; NID:935517; PIDN:CAA30564.1; PID:935518
A:Note: 92-Arg was also found
C:Superfamily: pleckstrin; pleckstrin repeat homology
C:Keywords: phosphoprotein; signal transduction
F:3-99/Domain: pleckstrin repeat homology <PLK1>
F:243-345/Domain: pleckstrin repeat homology <PLK2>

S00755 Length: 350 February 11, 2000 15:51 Type: P Check: 1536

1 MEPRRIEY LVKGSVENT WKPMVYLLE DGIEFYKKKS DNSPKGMIPL
51 KGSTLTPQC DCKRMFVK ITTKQODH FQAALFEERD AWWDINKAI
101 KCIEGOKFA RKSTRSIRL PETIDGALV LSKMOTERGI KELULEDKK
151 IFNCTFGNC VIDMLVSNQ VNRNQCIML NSLLNEGL QPAGDKSKSA
201 VDGTAENPFL DNPDAFYFP DSGFCEENS SDDVILKEE FRGIIKQC
251 LKQGHRRKN WKYKFIKRE DPAVLIHYDP AGADEPLGAI HLRCCVTSV
301 ESNNGKSE EENLEIITA DEVHYFLQAA TKRENTENIK AIQWASRTGK

!!AA_SEQUENCE 1.0
P1:HLH06 - class II histocompatibility antigen-associated gamma chain - human
N:Alternate names: CD74 antigen; HLA-DR antigen-associated invariant chain; MHC
class II-associated invariant chain
C:Species: Homo sapiens (man)

```

C>Date: 04-Dec-1986 #sequence_revision 26-Jan-1996 #text_change 22-Jun-1999
 C/Accession: A93981; B93981; A93602; A94103; S28903; S07182; A33234; A02243;
 A27551; A30060; S46255
 R/Claesson, L.; Larhammar, D.; Rask, L.; Peterson, P.A.
 Proc. Natl. Acad. Sci. U.S.A. 80, 7395-7399, 1983
 A/Title: cDNA clone for the human invariant gamma chain of class II
 histocompatibility antigens and its implications for the protein structure.
 A/Reference number: A93981; MUID:84170234
 A/Molecule type: mRNA
 A/Accession: A93981
 A/Residues: 1-216 <CL1>
 A/Cross-references: GB:K01144; NID:9188469; PIDN:AAA6304.1; PID:9188470
 A/Accession: B93981
 A/Molecule type: protein
 A/Residues: 1,'XXXXX','7','XXXXX','14','X','16-17 <CL2>
 A/Note: radiochemical amino acid sequencing after cell-free translation
 R/Kudo, U.; Chao, L.Y.; Naito, E.; Saunders, G.F.
 Nucleic Acids Res. 13, 8827-8841, 1985
 A/Title: Structure of the human gene encoding the invariant gamma-chain of
 class II histocompatibility antigens.
 A/Reference number: A93602; MUID:86093681
 A/Accession: A93602
 A/Molecule type: DNA
 A/Residues: 'MHRRRSRSCREDQPV',1-150,'T',152-216 <KUD>
 A/Note: the authors translated the codon CAG for residue 109 as Gly, AAG for
 residue 143 as Gln, and ACA for residue 151 as Arg
 R/O'Sullivan, D.M.; Larhammar, D.; Wilson, M.C.; Peterson, P.A.; Quaranta, V.
 Proc. Natl. Acad. Sci. U.S.A. 83, 4484-4488, 1986
 A/Title: Structure of the human Ia-associated invariant (gamma)-chain gene:
 identification of 5' sequences shared with major histocompatibility complex
 class II genes.
 A/Reference number: A94103; MUID:86233451
 A/Accession: A94103
 A/Molecule type: DNA
 A/Residues: 1-216 <OSU>
 R/Riberty, J.M.; Newcomb, J.R.; Surman, M.J.; Barbosa, J.A.; Cresswell, P.
 Nature 360, 474-477, 1992
 A/Title: HLA-DR molecules from an antigen-processing mutant cell line are
 associated with invariant chain peptides.
 A/Reference number: S28903; MUID:93078879
 A/Accession: S28903
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 81-104 <RIB>
 R/Strubin, M.; Mach, B.; Long, E.O.
 EMBO J. 3, 869-872, 1984
 A/Title: The complete sequence of the mRNA for the HLA-DR-associated invariant
 chain reveals a polypeptide with an unusual transmembrane polarity.
 A/Reference number: S07182; MUID:84207945
 A/Accession: S07182
 A/Molecule type: mRNA
 A/Residues: 'MHRRRSRSCREDQPV',1-216 <STR>
 A/Cross-references: EMBL:X00447; NID:932130; PIDN:CAA25192.1; PID:932131
 A/Accession: A33234
 A/Molecule type: DNA
 A/Residues: 'MHRRRSRSCREDQPV',1-4 <STR2>
 R/Katsumata, N.; Kakegawa, H.; Matsunaga, Y.; Saitara, T.
 FEBS Lett. 349, 265-269, 1994
 A/Title: Immunological significances of invariant chain from the aspect of its
 structural homology with the cystatin family.
 A/Reference number: S46255; MUID:94326933
 A/Accession: S46255
 A/Contents: annotation
 A/Note: some conclusions in this reference are based on the assumption, with no
 experimental evidence, that an upstream initiator is used
 C/Comment: Class II antigens are associated with the invariant gamma chain
 during intracellular transport. In the Golgi apparatus, they are glycosylated
 and dissociate from the gamma chain. Some gamma chains become independently
 integrated into the plasma membrane.
 C/Comment: Both cell-free translation experiments and the consensus sequence
 for the initiator of translation (Kozak) suggest that Met-1 is the initiator
 codon. Some authors have shown the translation of an additional sixteen
 residues upstream of the initiator, but this region is not likely to expressed.
 C/Genetics:

A/Gene: GDB:CD74; D1LAC
 A/Cross-references: GDB:119846; OMIM:142790
 A/Map position: 5q31-5q33
 A/Intons: 26/2; 84/1; 110/3; 131/3; 163/3; 193/1; 214/1
 C/Superfamily: class II histocompatibility antigen-associated gamma chain;
 thyroglobulin type I repeat homology
 C/Keywords: glycoprotein; transmembrane protein
 F:1-30/Domain: intracellular #status predicted <INT>
 F:31-56/Domain: transmembrane #status predicted <TM>
 F:57-216/Domain: extracellular #status predicted <EXT>
 F:114,120/Binding site: carbohydrate (Asn) (covalent) #status predicted

HMUG Length: 216 February 11, 2000 15:51 Type: P Check: 908 ..

1 MDOQRDLISN NEOLPMIGRR PGAFESKCR GALYGFSL VTLILAQAT
 51 TATFLIQOQG RLDKLTIVSQ NLOLENIAM LKRPKPVSK MRNATPLMQ
 101 ALPMGALPQG PQNATKRYGN MTEDEVNHL QNADPLKVP PLKSPFENL
 151 RHKNITWETI DWKFEESWNR HWLFENSRH SLEQKPTDAP PRESLEDEP
 201 SSGIGVTKOD LGPVP

11AA_SEQUENCE 1.0
 P1:R31V8 - ribosomal protein S8 - liverwort (Marchantia polymorpha) chloroplast
 C/Species: chloroplast Marchantia polymorpha
 C/Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 22-Jun-1999
 C/Accession: A02716; S01561
 R/Ohyama, K.
 submitted to the EMBL Data Library, October 1986
 A/Reference number: A00150
 A/Accession: A02716
 A/Molecule type: DNA
 A/Residues: 1-132 <OHY>
 R/Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.;
 Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z.; Aota, S.; Inokuchi, H.; Ozeki,
 H.
 Nature 322, 572-574, 1986
 A/Title: Chloroplast gene organization deduced from complete sequence of
 liverwort Marchantia polymorpha chloroplast DNA.
 A/Reference number: A38014
 A/Contents: annotation; gene organization; sites; features
 R/Fukuzawa, H.; Kohchi, T.; Sano, T.; Shirai, H.; Umesono, K.; Inokuchi, H.;
 Ozeki, H.; Ohyama, K.
 J. Mol. Biol. 203, 333-351, 1988
 A/Title: Structure and organization of Marchantia polymorpha chloroplast
 genome. III. Gene organization of the large single copy region from rbcL to
 trnI(CAU).
 A/Reference number: S01529; MUID:89068687
 A/Accession: S01561
 A/Molecule type: DNA
 A/Residues: 1-132 <OHY>
 A/Cross-references: GB:X04465; GB:Y00686; NID:911640; PIDN:CAA28121.1;
 PID:911710
 C/Genetics:
 A/Gene: rps8
 A/Gene: rps8
 A/Gene: chloroplast
 C/Superfamily: Escherichia coli ribosomal protein S8
 C/Keywords: chloroplast; protein biosynthesis; ribosome

R31V8 Length: 132 February 11, 2000 15:51 Type: P Check: 6002 ..

1 MGNDTIANMI TSIRNANLGR IKTVQVATN ITRNIAKILF QEGFIDNFID
 51 NKONTDILL INLKYGKRR KSYITLRR SRPGLRYSN HKRIPKVLGG
 101 MGVILSTR GIMTREAQ KKIIGELLCY VM

11AA_SEQUENCE 1.0
 P1:S2696 - ribosomal protein S13.e. cytosolic - fission yeast
 (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe

C>Date: 29-Jan-1993 #sequence-revision 19-Apr-1996 #text-change 22-Jun-1999
 C/Accession: S26296
 R/Marks: J.; Simanis, V.
 Nucleic Acids Res. 20, 4094, 1992
 A/Title: Cloning of the gene for ribosomal protein S13 from the fission yeast
 Schizosaccharomyces pombe.
 A/Reference number: S26296; MUID:92375702
 A/Accession: S26296
 A/Molecule type: DNA
 A/Residues: 1-151 <MAR>
 A/Cross-references: EMBL:X67030; NID:95071; PIDN:CAA47424.1; PID:95072
 C/Genetics:
 C/Introns: 68/1: 79/1: 94/3
 C/Superfamily: rat ribosomal protein S13; eubacterial ribosomal protein S15
 homology
 C/Keywords: protein biosynthesis; ribosome
 F/2-151/Product: ribosomal protein S13.e #status predicted <MAT>
 F/82-148/Domain: eubacterial ribosomal protein S15 homology <S15>
 S26296 Length: 151 February 11, 2000 15:51 Type: P Check: 6968 ..

1 MGRHSHKSGK IASGALPYVR SPPAWCKADA DSVVEQILKF SKRGMSPSQI
 51 GVLDSHGI POVFTITGQR IMRIKANGL APELPEDLYN LIRKAVSYRK
 101 HLERNRKDKD SKFRLLILIES RIHRLARYR KVGALPPTWK YESATASALV
 151 A

1!AA-SEQUENCE 1.0
 P1:A48103 - ribosomal protein L21 precursor, chloroplast - spinach
 C/Species: Spinacia oleracea (spinach)
 C/Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999
 C/Accession: A48103; S13527; A37906
 R/Lagrange, T.; Franca, B.; Ayala, M.; Maché, R.; Leber-Maché, S.
 Mol. Cell. Biol. 13, 2614-2622, 1993
 A/Title: Structure and expression of the nuclear gene coding for the
 chloroplast ribosomal protein L21: developmental regulation of a housekeeping
 gene by alternative promoters
 A/Reference number: A48103; MUID:93205007
 A/Accession: A48103
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-256 <LAG>
 A/Cross-references: GB:M64682; NID:9310589; PIDN:AAA74715.1; PID:9310590
 A/Note: Sequence extracted from NCBI backbone (NCBIN:127727, NCBI:P:127728)
 R/Martin, W.; Lagrange, T.; Li, Y.F.; Bisanz-Seyer, C.; Maché, R.
 Curr. Genet. 18, 553-556, 1990
 A/Title: Hypothesis for the evolutionary origin of the chloroplast ribosomal
 protein L21 of spinach.
 A/Reference number: S13527; MUID:91168304
 A/Accession: S13527
 A/Molecule type: mRNA
 A/Residues: 1-256 <MAR>
 A/Cross-references: EMBL:X56691; NID:921314; PIDN:CAA40019.1; PID:921315
 R/Smock, P.M.; Kruff, V.; Subramanian, A.R.
 J. Biol. Chem. 265, 16699-16703, 1990
 A/Title: A ribosomal protein is encoded in the chloroplast DNA in a lower plant
 but in the nucleus in angiosperms. Isolation of the spinach L21 protein and
 cDNA clone with transit and an unusual repeat sequence.
 A/Reference number: A37906; MUID:90375547
 A/Accession: A37906
 A/Molecule type: mRNA
 A/Residues: 1-256 <SMO>
 A/Cross-references: GB:M57413; GB:M31763; NID:9170134; PIDN:AAA34041.1;
 PID:9170135
 C/Superfamily: Escherichia coli ribosomal protein L21
 C/Keywords: chloroplast; protein biosynthesis; ribosome
 F/1-53/Domain: transit peptide (chloroplast) #status predicted <TMP>
 F/36-256/Product: ribosomal protein L21 #status predicted <MAT>
 A48103 Length: 256 February 11, 2000 15:51 Type: P Check: 718 ..

1 MASATLAFSC SSLCATKLKP QNINPILNV PPLSKPFCV VSPPSISRLS
 51 LRVAAKRRR FOEIPBEIKA EEEFOPRRN QKPQSLDVLV DDFQADPEPT
 101 PEVNDINDINQ LPKKGPPEPR EEIFAVVIG SRQIVIVGR WIYTORLKA
 151 TVNDKIVLAK VLVGTRAST YIGTPIVTA AVHAWVEQL LDKVIVEXY
 201 KKKKRNRI GHROPITRIK ITGITGEDY PASTLEAVE AKKEAEAE
 251 AEAPV

1!AA-SEQUENCE 1.0
 P1:J01298 - annexin II type 2 - African clawed frog
 N/Alternate names: calpactin II; lipocortin II; p36
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 31-Mar-1992 #sequence-revision 26-May-1994 #text-change 22-Jun-1999
 C/Accession: J01298; A41002; B41002
 R/Geake, V.; Koch, W.; Thiel, C.
 Gene 104, 259-264, 1991
 A/Title: Primary structure and expression of the Xenopus laevis gene encoding
 annexin II.
 A/Reference number: J01297; MUID:92009222
 A/Accession: J01298
 A/Molecule type: mRNA
 A/Residues: 1-340 <GR>
 A/Cross-references: GB:M60769; NID:9214532; PIDN:AAA49886.1; PID:9214533
 A/Experimental source: Kidney
 R/Zant, U.G.; Bryson, L.J.
 U. Biol. Chem. 266, 18560-18566, 1991
 A/Title: Xenopus annexin II (calpactin I) heavy chain has a distinct amino
 terminus.
 A/Reference number: A41002; MUID:92011609
 A/Accession: A41002
 A/Molecule type: mRNA
 A/Residues: 1-222, 'HP', 225-296, 'RN', 299-340 <I2A>
 A/Cross-references: GB:M58575; NID:9214007; PIDN:AAA49664.1; PID:9214008;
 GB:M58576; NID:9214009; PID:9214010
 A/Experimental source: ovary; clones A3 and C4
 A/Accession: B41002
 A/Molecule type: mRNA
 A/Residues: 52-222, 'HP', 225-243, 'R', 245-257, 'S', 259-296, 'RN', 299-340 <I2Z>
 A/Cross-references: GB:M58577
 A/Experimental source: ovary; clones E4 and F4
 A/Note: translation of the nucleotide sequence is not complete
 C/Genetics:
 A/Gene: Anx II-2
 C/Superfamily: annexin I; annexin repeat homology
 C/Keywords: calcium binding; duplication; endonexin fold; inflammation;
 phospholipid binding; phosphoprotein
 F/2-340/Product: annexin II type 2 #status predicted <MAT>
 F/37-108/Domain: annexin repeat homology <AX1>
 F/18-64/Region: endonexin fold #status predicted
 F/109-186/Domain: annexin repeat homology <AX2>
 F/120-138/Region: endonexin fold #status predicted
 F/193-265/Domain: annexin repeat homology <AX3>
 F/205-221/Region: endonexin fold #status predicted
 F/269-340/Domain: annexin repeat homology <AX4>
 F/280-296/Region: endonexin fold #status predicted
 F/27/Binding site: phosphate (Ser) (covalent) #status predicted
 J01298 Length: 340 February 11, 2000 15:51 Type: P Check: 409 ..

1 MALIHEIGK LSLGNOSS ROSKUGSYKA ATHFDEKDA AAIETAKTK
 51 GVDELTINI LTRNSNQRO DIAFAFHRT KROLPSALG ALSGNLEIV
 101 LGLIKTRQY DASELAKMK GLGDEPDLI EICSTNKE LLDONNAYRE
 151 LKTELEKDI MSDTSGDFK LMAVALAGR QEDGNVDE KIIDDAEELY
 201 EAGVKKRGTD VTKWITIME RSISHLOKV ERKTSISPD IESIKREKV

251 GDLNAPLNL VOCIQNKPLV FADRLYESMK GKTKRDKILI RIMVSRCELD
301 MKRIQEFKK KYGKSLHYFI GQDTKGDYOR ALLNLGCGDD
!!AA_SEQUENCE 1.0
P1:OOF - rhodopsin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 22-Jun-1999
C:Accession: A90864; A90865; A22012
R:O'Touss, J.E.; Baehr, W.; Martin, R.L.; Hirsch, J.; Pak, W.L.; Applebury, M.L.
Cell 40, 839-850, 1985
A:Title: The Drosophila rhodopsin gene encodes an opsin.
A:Reference number: A90864; MUID:85176937
A:Accession: A90864
A:Molecule type: DNA
A:Residues: 1-373 <OTD>
A:Cross-references: GB:K02315; NID:g158007; PIDN:AAA28733.1; PID:g158008
R:Zuker, C.S.; Cowan, A.F.; Rubin, G.M.
Cell 40, 851-858, 1985
A:Title: Isolation and structure of a rhodopsin gene from Drosophila melanogaster.
A:Reference number: A90865; MUID:85176938
A:Accession: A90865
A:Molecule type: mRNA
A:Residues: 1-373 <ZUK>
C:Comment: The domains were proposed from hydrophathy indices.
C:Comment: Some or all of the carboxyl-terminal Ser or Thr residues may be phosphorylated.
C:Genetics:
A:Gene: rhodopsin
A:Cross-references: Flybase:FBgn0002940
A:Map position: 3866 (92B8-11)
A:Introns: 3/2; 190/2; 239/3; 332/2
C:Superfamily: vertebrate rhodopsin
C:Keywords: chromoprotein; G protein-coupled receptor; glycoprotein;
photoreceptor; retinal; transmembrane protein
F:1-49/Domain: extracellular #status predicted <EX1>
F:50-74/Domain: transmembrane #status predicted <TM1>
F:75-86/Domain: transmembrane #status predicted <TM2>
F:87-109/Domain: transmembrane #status predicted <TM3>
F:110-127/Domain: extracellular #status predicted <EX2>
F:128-153/Domain: transmembrane #status predicted <TM4>
F:154-160/Domain: transmembrane #status predicted <TM5>
F:161-181/Domain: transmembrane #status predicted <TM6>
F:182-215/Domain: extracellular #status predicted <EX3>
F:216-243/Domain: transmembrane #status predicted <TM7>
F:244-276/Domain: transmembrane #status predicted <TM8>
F:277-300/Domain: intracellular #status predicted <IN1>
F:301-332/Domain: transmembrane #status predicted <TM9>
F:333-337/Domain: extracellular #status predicted <EX4>
F:338-363/Domain: transmembrane #status predicted <TM10>
F:364-399/Domain: intracellular #status predicted <IN2>
F:400-435/Domain: transmembrane #status predicted <TM11>
F:436-471/Domain: extracellular #status predicted <EX5>
F:472-507/Domain: transmembrane #status predicted <TM12>
F:508-543/Domain: extracellular #status predicted <EX6>
F:544-579/Domain: transmembrane #status predicted <TM13>
F:580-615/Domain: extracellular #status predicted <EX7>
F:616-651/Domain: transmembrane #status predicted <TM14>
F:652-687/Domain: extracellular #status predicted <EX8>
F:688-723/Domain: transmembrane #status predicted <TM15>
F:724-759/Domain: extracellular #status predicted <EX9>
F:760-795/Domain: transmembrane #status predicted <TM16>
F:796-831/Domain: extracellular #status predicted <EX10>
F:832-867/Domain: transmembrane #status predicted <TM17>
F:868-903/Domain: extracellular #status predicted <EX11>
F:904-939/Domain: transmembrane #status predicted <TM18>
F:940-975/Domain: extracellular #status predicted <EX12>
F:976-1011/Domain: transmembrane #status predicted <TM19>
F:1012-1047/Domain: extracellular #status predicted <EX13>
F:1048-1083/Domain: transmembrane #status predicted <TM20>
F:1084-1119/Domain: extracellular #status predicted <EX14>
F:1120-1155/Domain: transmembrane #status predicted <TM21>
F:1156-1191/Domain: extracellular #status predicted <EX15>
F:1192-1227/Domain: transmembrane #status predicted <TM22>
F:1228-1263/Domain: extracellular #status predicted <EX16>
F:1264-1299/Domain: transmembrane #status predicted <TM23>
F:1300-1335/Domain: extracellular #status predicted <EX17>
F:1336-1371/Domain: transmembrane #status predicted <TM24>
F:1372-1407/Domain: extracellular #status predicted <EX18>
F:1408-1443/Domain: transmembrane #status predicted <TM25>
F:1444-1479/Domain: extracellular #status predicted <EX19>
F:1480-1515/Domain: transmembrane #status predicted <TM26>
F:1516-1551/Domain: extracellular #status predicted <EX20>
F:1552-1587/Domain: transmembrane #status predicted <TM27>
F:1588-1623/Domain: extracellular #status predicted <EX21>
F:1624-1659/Domain: transmembrane #status predicted <TM28>
F:1660-1695/Domain: extracellular #status predicted <EX22>
F:1696-1731/Domain: transmembrane #status predicted <TM29>
F:1732-1767/Domain: extracellular #status predicted <EX23>
F:1768-1803/Domain: transmembrane #status predicted <TM30>
F:1804-1839/Domain: extracellular #status predicted <EX24>
F:1840-1875/Domain: transmembrane #status predicted <TM31>
F:1876-1911/Domain: extracellular #status predicted <EX25>
F:1912-1947/Domain: transmembrane #status predicted <TM32>
F:1948-1983/Domain: extracellular #status predicted <EX26>
F:1984-2019/Domain: transmembrane #status predicted <TM33>
F:2020-2055/Domain: extracellular #status predicted <EX27>
F:2056-2091/Domain: transmembrane #status predicted <TM34>
F:2092-2127/Domain: extracellular #status predicted <EX28>
F:2128-2163/Domain: transmembrane #status predicted <TM35>
F:2164-2199/Domain: extracellular #status predicted <EX29>
F:2200-2235/Domain: transmembrane #status predicted <TM36>
F:2236-2271/Domain: extracellular #status predicted <EX30>
F:2272-2307/Domain: transmembrane #status predicted <TM37>
F:2308-2343/Domain: extracellular #status predicted <EX31>
F:2344-2379/Domain: transmembrane #status predicted <TM38>
F:2380-2415/Domain: extracellular #status predicted <EX32>
F:2416-2451/Domain: transmembrane #status predicted <TM39>
F:2452-2487/Domain: extracellular #status predicted <EX33>
F:2488-2523/Domain: transmembrane #status predicted <TM40>
F:2524-2559/Domain: extracellular #status predicted <EX34>
F:2560-2595/Domain: transmembrane #status predicted <TM41>
F:2596-2631/Domain: extracellular #status predicted <EX35>
F:2632-2667/Domain: transmembrane #status predicted <TM42>
F:2668-2703/Domain: extracellular #status predicted <EX36>
F:2704-2739/Domain: transmembrane #status predicted <TM43>
F:2740-2775/Domain: extracellular #status predicted <EX37>
F:2776-2811/Domain: transmembrane #status predicted <TM44>
F:2812-2847/Domain: extracellular #status predicted <EX38>
F:2848-2883/Domain: transmembrane #status predicted <TM45>
F:2884-2919/Domain: extracellular #status predicted <EX39>
F:2920-2955/Domain: transmembrane #status predicted <TM46>
F:2956-2991/Domain: extracellular #status predicted <EX40>
F:2992-3027/Domain: transmembrane #status predicted <TM47>
F:3028-3063/Domain: extracellular #status predicted <EX41>
F:3064-3099/Domain: transmembrane #status predicted <TM48>
F:3100-3135/Domain: extracellular #status predicted <EX42>
F:3136-3171/Domain: transmembrane #status predicted <TM49>
F:3172-3207/Domain: extracellular #status predicted <EX43>
F:3208-3243/Domain: transmembrane #status predicted <TM50>
F:3244-3279/Domain: extracellular #status predicted <EX44>
F:3280-3315/Domain: transmembrane #status predicted <TM51>
F:3316-3351/Domain: extracellular #status predicted <EX45>
F:3352-3387/Domain: transmembrane #status predicted <TM52>
F:3388-3423/Domain: extracellular #status predicted <EX46>
F:3424-3459/Domain: transmembrane #status predicted <TM53>
F:3460-3495/Domain: extracellular #status predicted <EX47>
F:3496-3531/Domain: transmembrane #status predicted <TM54>
F:3532-3567/Domain: extracellular #status predicted <EX48>
F:3568-3603/Domain: transmembrane #status predicted <TM55>
F:3604-3639/Domain: extracellular #status predicted <EX49>
F:3640-3675/Domain: transmembrane #status predicted <TM56>
F:3676-3711/Domain: extracellular #status predicted <EX50>
F:3712-3747/Domain: transmembrane #status predicted <TM57>
F:3748-3783/Domain: extracellular #status predicted <EX51>
F:3784-3819/Domain: transmembrane #status predicted <TM58>
F:3820-3855/Domain: extracellular #status predicted <EX52>
F:3856-3891/Domain: transmembrane #status predicted <TM59>
F:3892-3927/Domain: extracellular #status predicted <EX53>
F:3928-3963/Domain: transmembrane #status predicted <TM60>
F:3964-3999/Domain: extracellular #status predicted <EX54>
F:4000-4035/Domain: transmembrane #status predicted <TM61>
F:4036-4071/Domain: extracellular #status predicted <EX55>
F:4072-4107/Domain: transmembrane #status predicted <TM62>
F:4108-4143/Domain: extracellular #status predicted <EX56>
F:4144-4179/Domain: transmembrane #status predicted <TM63>
F:4180-4215/Domain: extracellular #status predicted <EX57>
F:4216-4251/Domain: transmembrane #status predicted <TM64>
F:4252-4287/Domain: extracellular #status predicted <EX58>
F:4288-4323/Domain: transmembrane #status predicted <TM65>
F:4324-4359/Domain: extracellular #status predicted <EX59>
F:4360-4395/Domain: transmembrane #status predicted <TM66>
F:4396-4431/Domain: extracellular #status predicted <EX60>
F:4432-4467/Domain: transmembrane #status predicted <TM67>
F:4468-4503/Domain: extracellular #status predicted <EX61>
F:4504-4539/Domain: transmembrane #status predicted <TM68>
F:4540-4575/Domain: extracellular #status predicted <EX62>
F:4576-4611/Domain: transmembrane #status predicted <TM69>
F:4612-4647/Domain: extracellular #status predicted <EX63>
F:4648-4683/Domain: transmembrane #status predicted <TM70>
F:4684-4719/Domain: extracellular #status predicted <EX64>
F:4720-4755/Domain: transmembrane #status predicted <TM71>
F:4756-4791/Domain: extracellular #status predicted <EX65>
F:4792-4827/Domain: transmembrane #status predicted <TM72>
F:4828-4863/Domain: extracellular #status predicted <EX66>
F:4864-4899/Domain: transmembrane #status predicted <TM73>
F:4900-4935/Domain: extracellular #status predicted <EX67>
F:4936-4971/Domain: transmembrane #status predicted <TM74>
F:4972-5007/Domain: extracellular #status predicted <EX68>
F:5008-5043/Domain: transmembrane #status predicted <TM75>
F:5044-5079/Domain: extracellular #status predicted <EX69>
F:5080-5115/Domain: transmembrane #status predicted <TM76>
F:5116-5151/Domain: extracellular #status predicted <EX70>
F:5152-5187/Domain: transmembrane #status predicted <TM77>
F:5188-5223/Domain: extracellular #status predicted <EX71>
F:5224-5259/Domain: transmembrane #status predicted <TM78>
F:5260-5295/Domain: extracellular #status predicted <EX72>
F:5296-5331/Domain: transmembrane #status predicted <TM79>
F:5332-5367/Domain: extracellular #status predicted <EX73>
F:5368-5403/Domain: transmembrane #status predicted <TM80>
F:5404-5439/Domain: extracellular #status predicted <EX74>
F:5440-5475/Domain: transmembrane #status predicted <TM81>
F:5476-5511/Domain: extracellular #status predicted <EX75>
F:5512-5547/Domain: transmembrane #status predicted <TM82>
F:5548-5583/Domain: extracellular #status predicted <EX76>
F:5584-5619/Domain: transmembrane #status predicted <TM83>
F:5620-5655/Domain: extracellular #status predicted <EX77>
F:5656-5691/Domain: transmembrane #status predicted <TM84>
F:5692-5727/Domain: extracellular #status predicted <EX78>
F:5728-5763/Domain: transmembrane #status predicted <TM85>
F:5764-5799/Domain: extracellular #status predicted <EX79>
F:5800-5835/Domain: transmembrane #status predicted <TM86>
F:5836-5871/Domain: extracellular #status predicted <EX80>
F:5872-5907/Domain: transmembrane #status predicted <TM87>
F:5908-5943/Domain: extracellular #status predicted <EX81>
F:5944-5979/Domain: transmembrane #status predicted <TM88>
F:5980-6015/Domain: extracellular #status predicted <EX82>
F:6016-6051/Domain: transmembrane #status predicted <TM89>
F:6052-6087/Domain: extracellular #status predicted <EX83>
F:6088-6123/Domain: transmembrane #status predicted <TM90>
F:6124-6159/Domain: extracellular #status predicted <EX84>
F:6160-6195/Domain: transmembrane #status predicted <TM91>
F:6196-6231/Domain: extracellular #status predicted <EX85>
F:6232-6267/Domain: transmembrane #status predicted <TM92>
F:6268-6303/Domain: extracellular #status predicted <EX86>
F:6304-6339/Domain: transmembrane #status predicted <TM93>
F:6340-6375/Domain: extracellular #status predicted <EX87>
F:6376-6411/Domain: transmembrane #status predicted <TM94>
F:6412-6447/Domain: extracellular #status predicted <EX88>
F:6448-6483/Domain: transmembrane #status predicted <TM95>
F:6484-6519/Domain: extracellular #status predicted <EX89>
F:6520-6555/Domain: transmembrane #status predicted <TM96>
F:6556-6591/Domain: extracellular #status predicted <EX90>
F:6592-6627/Domain: transmembrane #status predicted <TM97>
F:6628-6663/Domain: extracellular #status predicted <EX91>
F:6664-6699/Domain: transmembrane #status predicted <TM98>
F:6700-6735/Domain: extracellular #status predicted <EX92>
F:6736-6771/Domain: transmembrane #status predicted <TM99>
F:6772-6807/Domain: extracellular #status predicted <EX93>
F:6808-6843/Domain: transmembrane #status predicted <TM100>
F:6844-6879/Domain: extracellular #status predicted <EX94>
F:6880-6915/Domain: transmembrane #status predicted <TM101>
F:6916-6951/Domain: extracellular #status predicted <EX95>
F:6952-6987/Domain: transmembrane #status predicted <TM102>
F:6988-7023/Domain: extracellular #status predicted <EX96>
F:7024-7059/Domain: transmembrane #status predicted <TM103>
F:7060-7095/Domain: extracellular #status predicted <EX97>
F:7096-7131/Domain: transmembrane #status predicted <TM104>
F:7132-7167/Domain: extracellular #status predicted <EX98>
F:7168-7203/Domain: transmembrane #status predicted <TM105>
F:7204-7239/Domain: extracellular #status predicted <EX99>
F:7240-7275/Domain: transmembrane #status predicted <TM106>
F:7276-7311/Domain: extracellular #status predicted <EX100>
F:7312-7347/Domain: transmembrane #status predicted <TM107>
F:7348-7383/Domain: extracellular #status predicted <EX101>
F:7384-7419/Domain: transmembrane #status predicted <TM108>
F:7420-7455/Domain: extracellular #status predicted <EX102>
F:7456-7491/Domain: transmembrane #status predicted <TM109>
F:7492-7527/Domain: extracellular #status predicted <EX103>
F:7528-7563/Domain: transmembrane #status predicted <TM110>
F:7564-7599/Domain: extracellular #status predicted <EX104>
F:7600-7635/Domain: transmembrane #status predicted <TM111>
F:7636-7671/Domain: extracellular #status predicted <EX105>
F:7672-7707/Domain: transmembrane #status predicted <TM112>
F:7708-7743/Domain: extracellular #status predicted <EX106>
F:7744-7779/Domain: transmembrane #status predicted <TM113>
F:7780-7815/Domain: extracellular #status predicted <EX107>
F:7816-7851/Domain: transmembrane #status predicted <TM114>
F:7852-7887/Domain: extracellular #status predicted <EX108>
F:7888-7923/Domain: transmembrane #status predicted <TM115>
F:7924-7959/Domain: extracellular #status predicted <EX109>
F:7960-7995/Domain: transmembrane #status predicted <TM116>
F:7996-8031/Domain: extracellular #status predicted <EX110>
F:8032-8067/Domain: transmembrane #status predicted <TM117>
F:8068-8103/Domain: extracellular #status predicted <EX111>
F:8104-8139/Domain: transmembrane #status predicted <TM118>
F:8140-8175/Domain: extracellular #status predicted <EX112>
F:8176-8211/Domain: transmembrane #status predicted <TM119>
F:8212-8247/Domain: extracellular #status predicted <EX113>
F:8248-8283/Domain: transmembrane #status predicted <TM120>
F:8284-8319/Domain: extracellular #status predicted <EX114>
F:8320-8355/Domain: transmembrane #status predicted <TM121>
F:8356-8391/Domain: extracellular #status predicted <EX115>
F:8392-8427/Domain: transmembrane #status predicted <TM122>
F:8428-8463/Domain: extracellular #status predicted <EX116>
F:8464-8499/Domain: transmembrane #status predicted <TM123>
F:8500-8535/Domain: extracellular #status predicted <EX117>
F:8536-8571/Domain: transmembrane #status predicted <TM124>
F:8572-8607/Domain: extracellular #status predicted <EX118>
F:8608-8643/Domain: transmembrane #status predicted <TM125>
F:8644-8679/Domain: extracellular #status predicted <EX119>
F:8680-8715/Domain: transmembrane #status predicted <TM126>
F:8716-8751/Domain: extracellular #status predicted <EX120>
F:8752-8787/Domain: transmembrane #status predicted <TM127>
F:8788-8823/Domain: extracellular #status predicted <EX121>
F:8824-8859/Domain: transmembrane #status predicted <TM128>
F:8860-8895/Domain: extracellular #status predicted <EX122>
F:8896-8931/Domain: transmembrane #status predicted <TM129>
F:8932-8967/Domain: extracellular #status predicted <EX123>
F:8968-9003/Domain: transmembrane #status predicted <TM130>
F:9004-9039/Domain: extracellular #status predicted <EX124>
F:9040-9075/Domain: transmembrane #status predicted <TM131>
F:9076-9111/Domain: extracellular #status predicted <EX125>
F:9112-9147/Domain: transmembrane #status predicted <TM132>
F:9148-9183/Domain: extracellular #status predicted <EX126>
F:9184-9219/Domain: transmembrane #status predicted <TM133>
F:9220-9255/Domain: extracellular #status predicted <EX127>
F:9256-9291/Domain: transmembrane #status predicted <TM134>
F:9292-9327/Domain: extracellular #status predicted <EX128>
F:9328-9363/Domain: transmembrane #status predicted <TM135>
F:9364-9399/Domain: extracellular #status predicted <EX129>
F:9400-9435/Domain: transmembrane #status predicted <TM136>
F:9436-9471/Domain: extracellular #status predicted <EX130>
F:9472-9507/Domain: transmembrane #status predicted <TM137>
F:9508-9543/Domain: extracellular #status predicted <EX131>
F:9544-9579/Domain: transmembrane #status predicted <TM138>
F:9580-9615/Domain: extracellular #status predicted <EX132>
F:9616-9651/Domain: transmembrane #status predicted <TM139>
F:9652-9687/Domain: extracellular #status predicted <EX133>
F:9688-9723/Domain: transmembrane #status predicted <TM140>
F:9724-9759/Domain: extracellular #status predicted <EX134>
F:9760-9795/Domain: transmembrane #status predicted <TM141>
F:9796-9831/Domain: extracellular #status predicted <EX135>
F:9832-9867/Domain: transmembrane #status predicted <TM142>
F:9868-9903/Domain: extracellular #status predicted <EX136>
F:9904-9939/Domain: transmembrane #status predicted <TM143>
F:9940-9975/Domain: extracellular #status predicted <EX137>
F:9976-10011/Domain: transmembrane #status predicted <TM144>
F:10012-10047/Domain: extracellular #status predicted <EX138>
F:10048-10083/Domain: transmembrane #status predicted <TM145>
F:10084-10119/Domain: extracellular #status predicted <EX139>
F:10120-10155/Domain: transmembrane #status predicted <TM146>
F:10156-10191/Domain: extracellular #status predicted <EX140>
F:10192-10227/Domain: transmembrane #status predicted <TM147>
F:10228-10263/Domain: extracellular #status predicted <EX141>
F:10264-10299/Domain: transmembrane #status predicted <TM148>
F:10300-10335/Domain: extracellular #status predicted <EX142>
F:10336-10371/Domain: transmembrane #status predicted <TM149>
F:10372-10407/Domain: extracellular #status predicted <EX143>
F:10408-10443/Domain: transmembrane #status predicted <TM150>
F:10444-10479/Domain: extracellular #status predicted <EX144>
F:10480-10515/Domain: transmembrane #status predicted <TM151>
F:10516-10551/Domain: extracellular #status predicted <EX145>
F:10552-10587/Domain: transmembrane #status predicted <TM152>
F:10588-10623/Domain: extracellular #status predicted <EX146>
F:10624-10659/Domain: transmembrane #status predicted <TM153>
F:10660-10695/Domain: extracellular #status predicted <EX147>
F:10696-10731/Domain: transmembrane #status predicted <TM154>
F:10732-10767/Domain: extracellular #status predicted <EX148>
F:10768-10803/Domain: transmembrane #status predicted <TM155>
F:10804-10839/Domain: extracellular #status predicted <EX149>
F:10840-10875/Domain: transmembrane #status predicted <TM156>
F:10876-10911/Domain: extracellular #status predicted <EX150>
F:10912-10947/Domain: transmembrane #status predicted <TM157>
F:10948-10983/Domain: extracellular #status predicted <EX151>
F:10984-11019/Domain: transmembrane #status predicted <TM158>
F:11020-11055/Domain: extracellular #status predicted <EX152>
F:11056-11091/Domain: transmembrane #status predicted <TM159>
F:11092-11127/Domain: extracellular #status predicted <EX153>
F:11128-11163/Domain: transmembrane #status predicted <TM160>
F:11164-11199/Domain: extracellular #status predicted <EX154>
F:11200-11235/Domain: transmembrane #status predicted <TM161>
F:11236-11271/Domain: extracellular #status predicted <EX155>
F:11272-11307/Domain: transmembrane #status predicted <TM162>
F:11308-11343/Domain: extracellular #status predicted <EX156>
F:11344-11379/Domain: transmembrane #status predicted <TM163>
F:11380-11415/Domain: extracellular #status predicted <EX157>
F:11416-11451/Domain: transmembrane #status predicted <TM164>
F:11452-11487/Domain: extracellular #status predicted <EX158>
F:11488-11523/Domain: transmembrane #status predicted <TM165>
F:11524-11559/Domain: extracellular #status predicted <EX159>
F:11560-11595/Domain: transmembrane #status predicted <TM166>
F:11596-11631/Domain: extracellular #status predicted <EX160>
F:11632-11667/Domain: transmembrane #status predicted <TM167>
F:11668-11703/Domain: extracellular #status predicted <EX161>
F:11704-11739/Domain: transmembrane #status predicted <TM168>
F:11740-11775/Domain: extracellular #status predicted <EX162>
F:11776-11811/Domain: transmembrane #status predicted <TM169>
F:11812-11847/Domain: extracellular #status predicted <EX163>
F:11848-11883/Domain: transmembrane #status predicted <TM170>
F:11884-11919/Domain: extracellular #status predicted <EX164>
F:11920-11955/Domain: transmembrane #status predicted <TM171>
F:11956-11991/Domain: extracellular #status predicted <EX165>
F:11992-12027/Domain: transmembrane #status predicted <TM172>
F:12028-12063/Domain: extracellular #status predicted <EX166>
F:12064-12099/Domain: transmembrane #status predicted <TM173>
F:12100-12135/Domain: extracellular #status predicted <EX167>
F:12136-12171/Domain: transmembrane #status predicted <TM174>
F:12172-12207/Domain: extracellular #status predicted <EX168>
F:12208-12243/Domain: transmembrane #status predicted <TM175>
F:12244-12279/Domain: extracellular #status predicted <EX169>
F:12280-12315/Domain: transmembrane #status predicted <TM176>
F:12316-12351/Domain: extracellular #status predicted <EX170>
F:12352-12387/Domain: transmembrane #status predicted <TM177>
F:12388-12423/Domain: extracellular #status predicted <EX171>
F:12424-12459/Domain: transmembrane #status predicted <TM178>
F:12460-12495/Domain: extracellular #status predicted <EX172>
F:12496-12531/Domain: transmembrane #status predicted <TM179>
F:12532-12567/Domain: extracellular #status predicted <EX173>
F:12568-12603/Domain: transmembrane #status predicted <TM180>
F:12604-12639/Domain: extracellular #status predicted <EX174>
F:12640-12675/Domain: transmembrane #status predicted <TM181>
F:12676-12711/Domain: extracellular #status predicted <EX175>
F:12712-12747/Domain: transmembrane #status predicted <TM182>
F:12748-12783/Domain: extracellular #status predicted <EX176>
F:12784-12819/Domain: transmembrane #status predicted <TM183>
F:12820-12855/Domain: extracellular #status predicted <EX177>
F:12856-12891/Domain: transmembrane #status predicted <TM184>
F:12892-12927/Domain: extracellular #status predicted <EX178>
F:12928-12963/Domain: transmembrane #status predicted <TM185>
F:12964-12999/Domain: extracellular #status predicted <EX179>
F:13000-13035/Domain: transmembrane #status predicted <TM186>
F:13036-13071/Domain: extracellular #status predicted <EX180>
F:13072-13107/Domain: transmembrane #status predicted <TM187>
F:13108-13143/Domain: extracellular #status predicted <EX181>
F:13144-13179/Domain: transmembrane #status predicted <TM188>
F:13180-13215/Domain: extracellular #status predicted <EX182>
F:13216-13251/Domain: transmembrane #status predicted <TM189>
F:13252-13287/Domain: extracellular #status predicted <EX183>
F:13288-13323/Domain: transmembrane #status predicted <TM190>
F:13324-13359/Domain: extracellular #status predicted <EX184>
F:13360-13395/Domain: transmembrane #status predicted <TM191>
F:13396-13431/Domain: extracellular #status predicted <EX185>
F:13432-13467/Domain: transmembrane #status predicted <TM192>
F:13468-13503/Domain: extracellular #status predicted <EX186>
F:13504-13539/Domain: transmembrane #status predicted <TM193>
F:13540-13575/Domain: extracellular #status predicted <EX187>
F

of the periplasmic binding protein-dependent transport system for peptides that are two to five amino acids long. This transport system is of importance to the nutrition of the organism; it is also essential for recycling of cell wall peptides.

C:Genetics:
A:Gene: oppB
A:Map position: 34 min
C:Function:
A:Description: probably responsible for mediating passage of peptides across the cytoplasmic membrane
C:Superfamily: oligopeptide permease protein oppB
C:Keywords: binding protein-dependent transport system; oligopeptide transport; transmembrane protein
F:102/Domain: transmembrane #status predicted <TM1>
F:103-121/Domain: transmembrane #status predicted <TM2>
F:140-156/Domain: transmembrane #status predicted <TM3>
F:173-189/Domain: transmembrane #status predicted <TM4>
F:228-244/Domain: transmembrane #status predicted <TM5>
F:273-289/Domain: transmembrane #status predicted <TM6>

OREBOB Length: 306 February 11, 2000 15:51 Type: P Check: 4920 ..

1 MKFTLRCL EAIPTLFLII TISFFMRLA PGSPTEGRA LPPEVLNIE
51 AKYHLNDPIM TOYFSYLKQL AHGDFGSEFK YKDYVNDLV AASEFVSACL
101 GAAMFLAVI IGVSAVIAA LKONTBMDYT VMGFAMTGVV IPSEFVAPLL
151 VMVFAITLQW LPGGWMNGA LKEMILPMVA LSLAYIASIA RITRGMIEV
201 LHSFIFTAR AKGLPMRRII FRHALKPALL PVLSTMGPAF VGITGSMVI
251 ETIYGLPGIG QLVFNGALNR DYSLVLSITI LVGALTILFN AIYDVLAVAI
301 DPKIRY

!!AA_SEQUENCE 1.0
P1:567566 - probable membrane protein YDL033c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein D2761
C:Species: *Saccharomyces cerevisiae*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A:Accession: S67566
R:Fullin, L.; Søren, A.M.; Laamanen, P.
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S67560
A:Accession: S67566
A:Molecule type: DNA
A:Residues: 1-417 <PAU>
A:Cross-references: EMBL:Z74081; NID:g1431011; PTDN:CAA98591.1; PID:e252967;
PID:g1431012; MIPS:YDL033c
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 4L
C:Superfamily: probable membrane protein YDL033c
C:Keywords: transmembrane protein
F:29-45/Domain: transmembrane #status predicted <TM>

S67566 Length: 417 February 11, 2000 15:51 Type: P Check: 8415 ..

1 MLARYNLIG RRSASPYRPO RLPAKFDNVI VAMSGVDS VAAALFAGEF
51 PNTGVMQON WSESQSLDP GKEPCYERDM RDVNRVAKHL NIRDVAVFE
101 ODYMDVFER MLRGYSEGST PNPDGCKNF VKFGKIREML DEKGTGVMY
151 LVYGHYARVM QEMNGKGLH LLRSIYRPD QSYLLSQINS TVLSILLAPI
201 GHTRKEPYRD LAKYAGLPTA EKPSQSGICE VNSQSGKRR NFLKHYLPSS
251 PGDITVDPO SGAKTWGRH DGLMSYITGG KVCISNPQAD PRYQGTWYVS
301 EKLRDTEIL IVRGDNPAL YSDTWRIENF SSIGPREDDI NAFONTGALT

351 LQRRSLQVPV QIKSCKLNRS ADNUDDTIHL ASKORATPBG QSCCLYIDR
401 VLGGSPISHV NNNDTHA

!!AA_SEQUENCE 1.0
P1:S30896 - virulence-associated protein mbal - *Salmonella choleraesuis* plasmid PKDSC50

C:Species: *Salmonella choleraesuis*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A:Accession: S30896; S20733
R:Matsu, H.; Abe, A.; Suzuki, S.; Kijima, M.; Tamura, Y.; Nakamura, M.; Kawahara, K.; Danbara, H.
Mol. Gen. Genet. 236, 219-226, 1993
A:Title: Molecular mechanism of the regulation of expression of plasmid-encoded mouse bacteremia (mba) genes in *Salmonella* serovar *Choleraesuis*.
A:Reference number: S30896; MUID:93173095
A:Accession: S30896
A:Molecule type: DNA
A:Residues: 1-297 <MAT>
A:Cross-references: EMBL:X54148; NID:g48765; PIDN:CAA38087.1; PID:g48766

C:Genetics:
A:Gene: mbal
C:Superfamily: virulence-associated protein spvr
C:Keywords: DNA binding, transcription regulation

S30896 Length: 297 February 11, 2000 15:51 Type: P Check: 7059 ..

1 MDPLINKLK IITLMEGTS FSATSVLYI TRTPLSRVIS DLRELKQRL
51 FIRNGTILIP TEPAOTIYRK VKSHYIFLHA LEOELGPCK TKOLEITFDE
101 IYPSLNLKI ISALTISGOK TNMRAVNS QIEELCQTN NCIVASRNY
151 FHRESIVCRT SVEGVMLEFI PKKEFLCGKP DINRLAGTVP LFHEGANFN
201 LDITVHFEQ TIGITNPAPS EDNVLEFSL YRLQGLANL LIPVYCRAL
251 GLSTDHALHI KGVALCTSLY YPRKRETPD YRKAKILQO ELKOSTE

!!AA_SEQUENCE 1.0
P1:S06670 - virulence-associated protein, mkaC - *Salmonella typhimurium* plasmid N:Alternate names: 33K virulence protein spvr
C:Species: *Salmonella typhimurium*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
A:Accession: S06670; A41318; S06089; S23714; S15213; C54540
R:Altra, S.; Chen, M.
EMBL:U01257; 274-278, 1989
A:Title: Molecular organization of genes constituting the virulence determinant on the *Salmonella typhimurium* 9c kilobase pair plasmid.
A:Reference number: S06670; MUID:90060335
A:Accession: S06670
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-297 <TAI>
A:Experimental source: virulence plasmid
R:Caldwell, A.L.; Guili, P.A.
J. Bacteriol. 173, 7176-7185, 1991
A:Title: The *Salmonella typhimurium* virulence plasmid encodes a positive regulator of a plasmid-encoded virulence gene.
A:Reference number: A41318; MUID:92041614
A:Accession: A41318
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <CAL>
A:Cross-references: GB:M74110
A:Experimental source: strain SR-11, virulence plasmid
R:Pullinger, G.D.; Baird, G.D.; Williamson, C.M.; Lax, A.J.
Nucleic Acids Res. 17, 7983, 1989
A:Title: Nucleotide sequence of a plasmid gene involved in the virulence of *Salmonella*.
A:Reference number: S06089; MUID:90016881
A:Accession: S06089

A:Molecule type: DNA
 A:Residues: 1-297 <POL>
 A:Cross-references: EMBL:X16111; NID:947859; PIDN:CAA34244.1; PID:947860
 A:Experimental source: isolate 1275 wild-type, virulence plasmid
 R:Note: F.; Pisanu, M.R.; Nicoli, J.; Popoff, M.Y.
 Res. Microbiol. 140, 627-630, 1989
 A>Title: A plasmid-borne virulence region (2.8 kb) from *Salmonella typhimurium* contains two open reading frames.
 A:Reference number: S23714; MUID:90176091
 A:Accession: S23714
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-297 <NOR>
 A:Cross-references: EMBL:X57092; NID:947788; PIDN:CAA40371.1; PID:947789
 A:Experimental source: strain C5, virulence plasmid
 R:Krusse, M.; Roudier, C.; Flarer, J.; Harwood, J.; Guiney, D.
 Mol. Microbiol. 5, 307-316, 1991
 A>Title: Molecular analysis of the virulence locus of the *Salmonella dublin* plasmid pSD2.
 A:Reference number: S15213; MUID:91251759
 A:Accession: S15213
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-40, 'G', 42-103, 'E', 105-297 <KRA>
 A:Cross-references: EMBL:X56727; NID:947836; PIDN:CAA40047.1; PID:947837
 R:Atara, S.; Baumann, M.; Rikonen, P.; Sukupolvi, S.; Rhen, M.
 FEMS Microbiol. Lett. 77, 319-324, 1991
 A>Title: Amino-terminal sequence analysis of four plasmid-encoded virulence-associated proteins of *Salmonella typhimurium*.
 A:Reference number: A54540
 A:Accession: C54540
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-10 <TA2>
 C:Genetics:
 A:Gene: mkac; mfc; spvr
 A:Genome: plasmid
 C:Superfamily: virulence-associated protein spvr
 C:KeyWords: DNA binding; transcription regulation
 S06670 Length: 297 February 11, 2000 15:51 Type: P Check: 7032

1 MDLIRKRLK IFTLMETGS FSIAISVLYI TRPLSRVYS DLREKQRL
 51 FIRKNGTLP TERAQIYRK VKSHYFLHA LEOEIGPTGK TKOLEIIFDE
 101 IYPSIKMLI ISALTISGQK TNIMGRAVNS QIEELCOTN NCIVISARNY
 151 FHRESIVCRT SVEGVMLFI PKFFLCGRP DINRLAGTPV LFHEGAKNFN
 201 LDTIHFEEQ TLGTTNPAFS FDNVDFSSL YRLQGGTAML LIVRRCRAL
 251 GLSTHAIHI KVALCTSLY YPTKKRTPD YRKAIKLIQ ELKQSF

!!AA:SEQUENCE 1.0
 P:IMCNA Colicin N immunity protein - *Escherichia coli* plasmid pCHAP4
 C:Species: *Escherichia coli*
 C:Date: 31-Mar-1991 #sequence, revision 31-Mar-1991 #text-change 16-Jul-1999
 C:Accession: S01761
 R:Pugsley, A.P.
 Mol. Gen. Genet. 211, 335-341, 1988
 A>Title: The immunity and lysis genes of *Colin* plasmid pCHAP4.
 A:Reference number: S01760; MUID:88174431
 A:Accession: S01761
 A:Molecule type: DNA
 A:Residues: 1-131 <PUG>
 A:Cross-references: EMBL:X06933; NID:941117; PIDN:CAA30020.1; PID:941119
 A:Note: the authors translated the codon ACC for residue 42 as Ile and GGT for residue 104 as Phe
 C:Genetics:
 A:Gene: cni
 A:Genome: plasmid
 C:Superfamily: colicin N immunity protein

C:Keywords: transmembrane protein
 F:66-84/Domain: transmembrane #status predicted <TM1>
 IMCNA Length: 131 February 11, 2000 15:51 Type: P Check: 6528

1 MHNTLEKII AYLSPFHS LNNPLSEAF NLVYHTPLA ATSLFTTRK
 51 ELKPKSSP LRAKILTPF TIIYISMTC FILDTELT SSXTEVLIYK
 101 KRSVFEPFL NTIYWDIYH IFVLVPRN I

!!AA:SEQUENCE 1.0
 P:ZMECD - proteinase umud (EC 3.4.21.-) - *Escherichia coli*
 N:Alternate names: umud; protease
 C:Species: *Escherichia coli*
 C:Date: 04-Dec-1986 #sequence, revision 04-Dec-1986 #text-change 16-Jul-1999
 C:Accession: A03551; R23157; D64864
 R:Kitagawa, Y.; Akaboshi, E.; Shinagawa, H.; Horii, T.; Ogawa, H.; Kato, T.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4356-4360, 1985
 A>Title: Structural analysis of the umu operon required for inducible mutagenesis in *Escherichia coli*.
 A:Reference number: A03550; MUID:85242679
 A:Accession: A03551
 A:Molecule type: DNA
 A:Residues: 1-139 <KIT>
 A:Cross-references: GB:M10107; NID:9148124; PIDN:AAA24728.1; PID:9148125
 R:Perry, K.L.; Ellledge, S.J.; Mitchell, B.B.; Marsh, L.; Walker, G.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4331-4335, 1985
 A>Title: umudC and umudB operators whose products are required for UV light- and chemical-induced mutagenesis: umud, umuA and umuB proteins share homology.
 A:Reference number: A23157; MUID:85242678
 A:Accession: A23157
 A:Molecule type: DNA
 A:Residues: 1-139 <PRY>
 A:Cross-references: GB:M13387; NID:9148127; PIDN:AAA98073.1; PID:9148128
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Klapetich, H.A.; Goeden, M.A.; Rose, D.O.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426517
 A:Accession: D64864
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-139 <BAT>
 A:Cross-references: GB:AE000216; GB:U00096; NID:91787417; PIDN:AACT4267.1; PID:91787431; UMG:P:1183
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: umud
 A:Map position: 26 min
 C:Complex: umud/umud homodimer; umud'/umud' homodimer; umud/umud' heterodimer
 C:Function:
 A:Description: essential for induced and SOS mutagenesis in *E. coli*; controlled by *RecA* and *LexA* gene products; self-cleavage of umud to umud' (mutagenetically active form); acts in concert with umudC, *RecA* and DNA polymerase III to facilitate the process of translesion synthesis, which results in the introduction of mutations
 A:Note: one of two proteins encoded by umu operon; intracellular levels of umud and umud' proteins are kept to a minimum by *lon* serine proteinase; umud' is removed from the cell by the *ClpXP* serine protease, but only when it is in a heterodimeric complex with umud
 C:Superfamily: LexA repressor
 C:Keywords: DNA repair; heterodimer; homodimer; hydrolase; induced mutagenesis; serine proteinase; SOS mutagenesis
 F:1-139/Product: umud protein #link #status predicted <PRO>
 F:1-24/Domain: propeptide #link #link #status predicted <PRO>
 F:25-139/Product: umud' protein #status predicted <MAT2>
 ZMECD Length: 139 February 11, 2000 15:51 Type: P Check: 8403

1 MLIKPADLR EIVTFPLFS LVQCFPSPA ADYVEKIDL NOLLIDPSA

51 TYFVKSAGDS MIDGDISDGD LLVDSNITA SHGDIVIAAV DGEFTVKKLO
 101 LRPVQOLIPM NSAVSPITIS SEPTLDVFGV VHVYKAKR

11AA_SEQUENCE 1.0
 P1:H64563 - flagellar motor switch protein - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: H64563
 R:Tomb, J.F.; White, O.; Kellavag, A.R.; Clayton, R.A.; Sutton, G.G.;
 Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
 Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
 B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenney, K.;
 Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne,
 J.D.; Uitterlbeck, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman,
 J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 Nature 388: 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
 Fraser, C.M.; Venter, J.C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter
 pylori.
 A:Accession: H64563
 A:Reference number: A64520; MWID:97394467
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1343 <TOM>
 A:Cross-references: GB:AE000552; GB:AE000511; NID:92313451; PIDN:AD07420.1;
 PID:92313453; TIGR:HP0352
 C:Superfamily: flagellar switch protein filig
 H64563 Length: 343 February 11, 2000 15:51 Type: P Check: 9038

1 MATKLTPEKOR ADELISMSE KIALLLIQVG EDTGELIRH LDDISTEIS
 51 KOIVOLNGTD KOIGAAVLEE FFAIFQSNQY INTGLEVAR ELLFRTLGSE
 101 EAKVMDKLT KSLQTOKNFA YLGKIKPOQL ADFIINERQ TIALILAME
 151 APNAAETLSY FPEMKAEIS IRMANLGEIS PQVYKRVSTV LENTLESITS
 201 YKIEVGGLRA VAEIFNRLQ KSAKTLARI ESDVNLAGA IKEMAFTEFD
 251 IVKLDNFAIR ELKVADKRD LSLAKTSTK DLTDFKLNW SSRAAEQFE
 301 EMQVLGAVKI KDYVDAQRI IEIVSLOEK GVITQGEED VIE

11AA_SEQUENCE 1.0
 P1:VEECU1 - unci protein - Escherichia coli
 A:Alternate names: NTP synthase subunit atpI
 C:Species: Escherichia coli
 C>Date: 02-Apr-1982 #sequence_revision 15-Oct-1982 #text_change 16-Jul-1999
 C:Accession: D30389; A23223; A34908; D65177; A04433
 R:Walker, J.E.; Gay, N.V.; Saraste, M.; Eberle, A.N.
 Biochem. J. 224: 799-815, 1984
 A:Title: DNA sequence around the Escherichia coli unc operon. Completion of the
 sequence of a 17 kilobase segment containing asna, orfC, unc, gims and phos.
 A:Reference number: A30389; MWID:85121806
 A:Accession: D30389
 A:Molecule type: DNA
 A:Residues: 1-130 <WAL>
 A:Cross-references: GB:X01631; NID:943256; PIDN:CAA25775.1; PID:9581248
 R:Kanazawa, H.; Kiyasu, T.; Noumi, T.; Futai, M.
 J. Bacteriol. 158: 300-306, 1984
 A:Title: Overproduction of subunit a of the F0 component of
 proton-translocation ATPase inhibits growth of Escherichia coli cells.
 A:Reference number: A23223; MWID:84185438
 A:Accession: A23223
 A:Molecule type: DNA
 A:Residues: 1-130 <KAN>
 R:Schnepp, B.; Deckers-Hebestreit, G.; Alendorff, K.
 J. Biol. Chem. 265: 389-395, 1990
 A:Title: Overproduction and purification of the unci gene product of the ATP
 synthase of Escherichia coli.

A:Reference number: A34908; MWID:90094427
 A:Accession: A34908
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-12 <SCH>
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277: 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MWID:97426617
 A:Accession: D65177
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-130 <BLAT>
 A:Cross-references: GB:AE000450; GB:U00096; NID:91790166; PIDN:AAC76762.1;
 PID:91790177; UWGP:b3739
 A:Experimental source: strain K-12, substrain MG1655
 R:Kanazawa, H.; Mabuuchi, K.; Futai, M.
 Biochem. Biophys. Res. Commun. 107: 568-575, 1982
 A:Title: Nucleotide sequence of the promoter region of the gene cluster for
 proton-translocating ATPase from Escherichia coli and identification of the
 active promoter.
 A:Reference number: I52197; MWID:83022404
 A:Accession: I52197
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 747-764 <RES>
 A:Cross-references: EMBL:V00313; NID:942286; PIDN:CAA23599.1; PID:9581172
 C:Comment: The unci gene is the first gene of the unc (or atp) operon. A
 possible function for this protein is to guide the assembly of the membrane
 sector of the proton transporting ATP synthase (EC 3.6.1.34) enzyme complex.
 C:Genetics:
 A:Gene: atpI; unci
 A:Map position: 84 min
 A:Start codon: Gtc
 C:Superfamily: unci protein
 C:Keywords: transmembrane protein
 VEECU1 Length: 130 February 11, 2000 15:51 Type: P Check: 4271

1 MKNWVSIVL SRNVAKKLL VOLLVIVASG LFLSLKPPW GVSASIGLA
 51 VFLPVLVEMI FAWRHOATP AKGRVAMTFA FGEAFKVLAM LVLLVALAV
 101 LKAVFLPLIV TWVLVLVQI LAPAVINNK

11AA_SEQUENCE 1.0
 P1:8AG55 - virB8 protein precursor - Agrobacterium tumefaciens plasmid pTi15955
 C:Species: Agrobacterium tumefaciens
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
 C:Accession: S00784
 R:Thompson, D.V.; Melchers, L.S.; Idler, K.B.; Schilperoort, R.A.; Hooykaas,
 P.J.T.
 Nucleic Acids Res. 16: 4621-4636, 1988
 A:Title: Analysis of the complete nucleotide sequence of the Agrobacterium
 tumefaciens virB operon
 A:Reference number: S00777; MWID:88247765
 A:Accession: S00784
 A:Molecule type: DNA
 A:Residues: 1-257 <TMO>
 A:Cross-references: EMBL:X06826; NID:939195; PIDN:CAA23978.1; PID:939210
 C:Genetics:
 A:Gene: plasmid
 C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein
 F/1-36/Domain: signal sequence #status predicted <SIG>
 F/37-257/Product: virB8 protein #status predicted <WAT>
 8AG55 Length: 257 February 11, 2000 15:51 Type: P Check: 173

1 MWGDSLLRQ IFSAIRVDA MTGPEYAMLV ARSLAEHKK EYEAFOIARA
 51 KSARLSKII AAVAAIILG NVAQAFATAT MPFLSLVAV YLMIRPDGT

101 DSEVSISRLP ATOBEAVVNA SIMEYVRLRE SYDADTAQYA YDLVSNFSAP
 151 TVRDYOOOF NYPNPSPOV ILGKRGRLEV EHIASNDVTP STQOIRYKRT
 201 LVDDKMPV STWTATVRE KVTSLPGRLR LTNPAVLVT SYQSEDTVS
 251 NVGQAGAP

!!AA_SEQUENCE 1.0
 P1:B8AG58 - virB8 protein - Agrobacterium tumefaciens plasmid pTiA6
 C:Species: Agrobacterium tumefaciens
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 03-Feb-1994
 C:Accession: J28621; J27127
 R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
 J. Biol. Chem. 265, 5804-5814, 1988
 A>Title: Characterization of the virB operon from an Agrobacterium tumefaciens
 Ti plasmid.
 A:Reference number: A28621; MUID:88186901
 A:Accession: J28621
 A:Molecule type: DNA
 A:Residues: 1-230 <MAR>
 A:Cross-references: GB:J03216
 A>Note: This sequence was designated ORF 9 in this reference
 R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.
 J. Biol. Chem. 265, 4768, 1990
 A:Reference number: A35737; MUID:90170994
 A:Contents: annotation; erratum
 C:Genetics:
 A:Genome: plasmid
 A:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein
 B8AG58 Length: 230 February 11, 2000 15:51 Type: P Check: 2646 ..

1 MIVAFESLAE HYKEVEAFQT ARKSKRRIS KIIAVAIA ILGNVACAPA
 51 IATWPLSLRL VPYVLMIRAD GTVDSEVSIS RLPAQGEAV VNASLWEIYR
 101 LRESDADTA QYAYDLVSNF SAEFTVADYQ QFENTPNSS PoyILGRGR
 151 VEVEHIASND VTPSTQOIRY KRLVVDGKM PVSWTATV RYKVTSLPG
 201 RLRLTNPAGL VVTSYQTSQD TVSNVQAGAP

!!AA_SEQUENCE 1.0
 P1:B8AG58 - virB8 protein - Agrobacterium tumefaciens plasmid pTiC58
 C:Species: Agrobacterium tumefaciens
 C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
 C:Accession: S12348; S1833; S10523
 R:Shirasu, K.; Morel, P.; Kado, C.I.
 Mol. Microbiol. 4, 1153-1163, 1990
 A>Title: Characterization of the virB operon of an Agrobacterium tumefaciens Ti
 plasmid: nucleotide sequence and protein analysis.
 A:Reference number: S12348; MUID:91041724
 A:Accession: S12348
 A:Molecule type: DNA
 A:Residues: 1-237 <ROG>
 A:Cross-references: EMBL:J03320; NID:9154781; PID:AAA91598.1; PID:9154790
 R:Rogowsky, P.M.; Powell, B.S.; Shirasu, K.; Lin, T.S.; Morel, P.; Zyprian,
 E.M.; Steck, T.R.; Kado, C.I.
 Plasmid 23, 85-106, 1990
 A>Title: Molecular characterization of the virB regulation of Agrobacterium
 tumefaciens: complete nucleotide sequence and gene organization of the
 28.63-kbp region cloned as a single unit.
 A:Reference number: S11825; MUID:90301800
 A:Accession: S11833
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-237 <ROG>
 A:Cross-references: EMBL:J03320; NID:9154781; PID:AAA91598.1; PID:9154790
 R:Kulda, G.A.; de Vos, G.; Owen, J.; McCaffrey, G.; Zambryski, P.
 Mol. Gen. Genet. 221, 256-266, 1990
 A>Title: The virB operon of Agrobacterium tumefaciens pTiC58 encodes 11 open

reading frames.
 A:Reference number: S10516; MUID:90318324
 A:Accession: S10523
 A:Molecule type: DNA
 A:Residues: 1-20,23-127, 'SA', 130-237 <KUL>
 A:Cross-references: EMBL:X53264; NID:939152; PID:CAA37361.1; PID:939160
 C:Genetics:
 A:Gene: virB8
 A:Genome: plasmid
 C:Superfamily: tumor-inducing plasmid pTiC58 virB8 protein
 B8AG58 Length: 237 February 11, 2000 15:51 Type: P Check: 6263 ..

1 MKGSEYALLV ARETLAEHYK EYEAFTQARA KSARLSKVI AAVATIAVLG
 51 NVQAQFTAT MVLPIRLVP YLMIRPDGV DSEVSISRLP ATOBEAVVNA
 101 SIMEYVRLRE SYDADTAQYA YDLVSNFTRP MYRONYQOOF NYPNPSPOV
 151 ILGKRGRLEV EHIASNDVTP GVOQIRYKRT LIVDGKMPMA STWTATVRE
 201 KVTSLPGRLR LTNPGGLVVT SYQSEDTVS NAGHSEP

!!AA_SEQUENCE 1.0
 P1:W2WL31 - E2 protein - human papillomavirus type 31
 C:Species: human papillomavirus type 31
 A:Note: host Homo sapiens (man)
 C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
 C:Accession: D32444
 R:Goldsbrough, M.D.; Disilvestre, D.; Temple, G.F.; Lorincz, A.T.
 Virology 179, 306-311, 1989
 A>Title: Nucleotide sequence of human papillomavirus type 31: a cervical
 neoplasia-associated virus.
 A:Reference number: A94398; MUID:89299478
 A:Accession: D32444
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-372 <GOL>
 A:Cross-references: GB:U04353; NID:9333048; PID:AAA6995.1; PID:9459919
 C:Superfamily: Papillomavirus E2 protein
 C:Keywords: DNA binding; early protein; transcription regulation
 W2WL31 Length: 372 February 11, 2000 15:51 Type: P Check: 9656 ..

1 METLSQRINV CODKLEHYE NDSKRLCDHI DYMKHRLKC VLMYKAREMG
 51 IHSINHOVVP ALSVSKAKAL QALEQOMME TLNTEYKNE DWTMOQTSLE
 101 LYLTPATGCL KKHGYVEVQ FGDVHNTWMH YTNKRFIYLC IDQOCTVVEG
 151 QVNCNGIYV HEGHIITYFN FTEBAKKYGT GKKEVHAGG QVAFPEVSF
 201 SDELSFPGI VTKLIPANNT TTSNSKTCAL GTSCEVRRAT TSKRRTPE
 251 EHRNTHFNRK LRGDSYDSV NCGVISAAC TNOTRAVSCP ATPPIHLKG
 301 DANILCLRY RLSKRYQLYE QVSTWHTC TDGKHNAIV TLTYISTSOR
 351 DFLNTVXIP NTVSVSTGYM TI

!!AA_SEQUENCE 1.0
 P1:W2WL31 - E2 protein - human papillomavirus type 51
 C:Species: human papillomavirus type 51
 A:Note: host Homo sapiens (man)
 C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Feb-1997
 C:Accession: B40415
 R:Lungu, O.; Crum, C.P.; Silverstein, S.J.
 J. Virol. 65, 4216-4225, 1991
 A>Title: Biologic properties and nucleotide sequence analysis of human
 papillomavirus type 51
 A:Reference number: A40415; MUID:91303675
 A:Accession: B40415
 A:Status: translation not shown

Mon Feb 14 08:07:32 2000

pir.cat

Page 61

```
A:Molecule type: DNA
A:Residues: 1-358 <LUN>
A:Cross-references: GB:M62877
C:Superfamily: papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
W2WL51 Length: 358 February 11, 2000 15:51 Type: P Check: 1816
1 METLCRLN COEKLIDCE LDCSKLVDO NWTLRLREA AMFYAREEN
51 LRTINQVAV ATTYSKQAC QAIMHVALO SLNKSDYNE PMWRETCYE
101 LMCVAKQCF KKGITVIV FDKNKNAMD YTSKFIITY DNDKRWKNG
151 NVDYGIYV VNSKEEYVQ FKDEAKIYGA QOMEVYMTGT VITCEPVSS
201 TCSDALSTT TVEQLSTPT TNPITVCVA KEAQOQKRR QRLTEPDSST
251 ISPLSVDNNT NQIHGSGST NTGCHOSATO TAFIYHLKGD TNCIKCFYR
301 FEHKGLYKN VSTWMTSN TKGTIVTVF DSAHQRETFI KTIKVPSTV
351 LSLGIWTL
!!A_SEQUENCE 1.0
P1:W2WL18 - E2 protein - human papillomavirus type 18
C:Species: human papillomavirus type 18
C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: D26251; S35634
R:Cole, S.T.; Danos, O.
J. Mol. Biol. 193, 599-608, 1987
A:Title: Nucleotide sequence and comparative analysis of the human
papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated
structure of the E6 and E7 gene products.
A:Reference number: A92937; MUID:87283862
A:Accession: D26251
A:Molecule type: DNA
A:Residues: 1-365 <COL>
A:Cross-references: GB:X05015; MID:960975; PIDN:CAA2667.1; PID:960979
R:Meisner, J.
Nucleic Acids Res. 21, 1041, 1993
A:Title: Tag1 is a single cut enzyme for HPV-18.
A:Reference number: S35634; MUID:93197132
A:Accession: S35634
A:Molecule type: DNA
A:Residues: 84-89, 'A', 91-93 <MEI>
A:Cross-references: EMBL:107917
C:Superfamily: papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
W2WL18 Length: 365 February 11, 2000 15:51 Type: P Check: 5567
1 MQPKETLSE RLSCVODKII DHYENDSKDI DSOIQWOLI RMENAFPPAA
51 REHQIOTLNL QVPAVINISK SKAKRALELO MALOGIACSR YKTEDWTLQD
101 TCEELWMTPE THCFKKGQT VOYVFDGNKD NCMTYVAMS VYWTDAQWT
151 DKATVCVSR GLYVKEGIN TTYIEFKSEC EKYGNNGTGE VHFNNVYDC
201 NDSMCSSTD TVSATOLVKO LOHTPSYSS TVSGVAKTY GQTSAAATPG
251 HGLAEKQRC GFVNPLGAA TPGNNKRRK LCGNTTPTI HIKGDRSLK
301 CLNRRLRKH DHYRDISSTW HWTGAGNEKT GILTYVHSE TORTRFLNTV
351 AIPDSVOILV GYMTM
!!A_SEQUENCE 1.0
P1:SL5617 - E2 protein - human papillomavirus type 2a
C:Species: human papillomavirus type 2a
A:Note: 1st Homo sapiens (man)
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 16-Feb-1997
C:Accession: S15617
R:Hirsch-Behnam, A.; Delius, H.; de Villiers, E.M.
Virus Res. 18, 81-98, 1990
A:Title: A comparative sequence analysis of two human papillomavirus (HPV)
types 2a and 57
A:Reference number: S15614; MUID:91188699
A:Accession: S15617
A:Molecule type: DNA
A:Residues: 1-391 <HR>
A:Cross-references: EMBL:X55964
C:Superfamily: papillomavirus E2 protein
C:Keywords: DNA binding; early protein; transcription regulation
S15617 Length: 391 February 11, 2000 15:51 Type: P Check: 734
1 METLANRLDA COETLLELXE KDSNKLEDO KHMAOVLLEN VMLERARECG
51 MTRVGTAVP ALTVSKARAC QALEVOLAO TLMQSAVSTE AMTLDTCLE
101 MMDAPPKCM KKGQSVLYK FDGSSDRDM YTSWGIIVQ DTIDSMHKV
151 PGVDELGLY YVHDGVNRY VDFGIESLTY GVTGWEVAV AGTVIHTTSA
201 SVSSTQASAS DDEPLSPIRT AVSPVPAPVA ASASSTGAGR AAPPTOALCS
251 AQAPTSPAK RQRIVGOQH PRPSTRTYG EGEVECYNKR SIDSNTDP
301 RMGHDIDSV PYIHLRGDAN CLKCFRRYQ KKKDVLXARY SSTWMAAGN
351 GDKTAEVTLW YTSVEQTEF LTRYSIPKGL IALPGYMSAF V
!!A_SEQUENCE 1.0
P1:QOBN5 - Integral membrane protein - simiirine herpesvirus 1 (strain 11)
C:Species: simiirine herpesvirus 1
A:Note: host simiiri solureus (common squirrel monkey)
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: A36810
R:Albrecht, J.
submitted to the EMBL Data Library, January 1992
A:Description: Primary structure of the herpesvirus simiiri genome.
A:Reference number: A36806
A:Accession: A36810
A:Molecule type: DNA
A:Residues: 1-366 <ALB>
A:Cross-references: GB:X64346; MID:960320; PIDN:CAA45662.1; PID:960360
R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.;
Newman, C.; Wiltmann, S.; Craxton, M.A.; Coleman, H.; Fleckenstein, B.; Honess,
R.W.
J. Virol. 66, 5047-5058, 1992
A:Title: Primary structure of the herpesvirus simiiri genome.
A:Reference number: A37309; MUID:92333688
A:Contents: annotation; possible protein-coding frames
A:Note: neither amino acid nor nucleotide sequence is given
C:Genetics:
A:Gene: 39
C:Superfamily: cytomagalovirus UL100 protein
C:Keywords: transmembrane protein
F:18-34/Domain: transmembrane #status predicted <TM>
F:87-103/Domain: transmembrane #status predicted <TM>
F:152-168/Domain: transmembrane #status predicted <TM>
F:214-230/Domain: transmembrane #status predicted <TM>
F:236-252/Domain: transmembrane #status predicted <TM>
F:274-290/Domain: transmembrane #status predicted <TM>
F:303-321/Domain: transmembrane #status predicted <TM>
QOBN5 Length: 366 February 11, 2000 15:51 Type: P Check: 7071
1 MKKASRSDTF MTRWIOILV LEVIMFMSA IIPIASVVG LGFCYFPNL
51 VDYLLWTL RNAAKHLPT LLEAPELFV YITWSVLVDL ASAIYVYVGA
101 LAILQARKTH LTSMTLQTM INVGSHTML FGIARWML OLPIHVSYSK
```

```
151  HVMALAEIYF LHFCLSYNMT LSLVSRNSPK WSVLLMEQHI PKOSLSTILL
201  DYKRPICVNM YLSLLALEML VFSLGPMMAI GNSFYIIVSD TVLASINLXF
251  VLTETWYMT EMFLQDYIYL QFGFYLGVS GSLILLFVL RYRAVVSAN
301  LKHTAVNIA MIPACVIAM MFLRFRSQO VKRPENSYTP LPRFRKRRQ
351  KQDQILME TSDEEL

11AA_SEQUENCE 1.0
P1:WMNZ22 Envelope-associated 22k protein - human respiratory syncytial virus
C:Species: human respiratory syncytial virus
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: B93010; A93009; A04034
R:Elango, N.; Satake, M.; Venkatesan, S.
J. Virol. 55, 101-110, 1985
A:Title: mRNA sequence of three respiratory syncytial virus genes encoding two
nonstructural proteins and a 22k structural protein.
A:Reference number: A93010; MUID:85237684
A:Accession: B93010
A:Molecule type: genomic RNA
A:Residues: 1-194 <ELAV>
A:Cross-references: GB:M11486; NID:933925; PIDN:AB59860.1; PID:933934
R:Collins, P.L.; Wertz, G.W.
J. Virol. 54, 65-71, 1985
A:Title: The envelope-associated 22k protein of human respiratory syncytial
virus: nucleotide sequence of the mRNA and a related polytranscript.
A:Reference number: A93009; MUID:85135082
A:Accession: A93009
A:Molecule type: genomic RNA
A:Residues: 1-194 <COL>
A:Cross-references: GB:M11486; GB:K01459; GB:K02719; GB:K03348; GB:K03349;
GB:M11217; GB:M1244; GB:M11487; GB:M11505; GB:M11514; GB:M11631; GB:M12966;
GB:X00001; GB:X02221; NID:933925; PIDN:AB59860.1; PID:933934
C:Genetics:
A:Gene: 22K
A:Superfamily: respiratory syncytial virus envelope-associated 22k protein
F:89,191/Binding site: carbohydrate (Asn) (covalent) #status predicted
WMNZ22 Length: 194 February 11, 2000 15:51 Type: P Check: 9828 ..

1 MSRRNPCKFE IRGHCNGKR CHFSHNYEWM PPHALLVRON FMNLRLKSM
101 KLTLENSDD IKRLDNEEL NSFIRVYNT VISLIESNRK NMQIHLK
151 RLPADVLKKT IKNTLDIHK IITNNPKST VSDINDIAKN NDTT

11AA_SEQUENCE 1.0
P1:WMNZBA - matrix glycoprotein M2 - bovine respiratory syncytial virus (strain
A151908)
C:Species: bovine respiratory syncytial virus
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: J01482
R:Zamorin, M.; Samal, S.K.
J. Gen. Virol. 73, 737-741, 1992
A:Title: Sequence analysis of M2 mRNA of bovine respiratory syncytial virus
obtained from an F-M2 distal oncolytic mRNA suggests structural homology with that
of human respiratory syncytial virus
A:Reference number: J01481; MUID:92185490
A:Accession: J01482
A:Molecule type: mRNA
A:Residues: 1-186 <ZAM>
A:Cross-references: GB:M82816; NID:9210823; PIDN:AAA42805.1; PID:9210825
C:Genetics:
A:Gene: M2
A:Superfamily: respiratory syncytial virus envelope-associated 22k protein
F:54,89/Binding site: carbohydrate (Asn) (covalent) #status predicted

WMNZBA Length: 186 February 11, 2000 15:51 Type: P Check: 6337 ..

1 MSRRNPCKFE IRGHCNGKR CHFSHNYEWM PPHALLVRON FMNLRLKSM
51 DRNNDLISEI SGAELDRTE EYALGVYVL ESYLSINNI TKOSACVAMS
101 KLALEINND IKRLNKEVP TSPKIRIYNT VISYDSNRK NMQIHLK
151 RLPADVLKKT IKNTIDINE INGNQGDIN VDEONE

11AA_SEQUENCE 1.0
P1:WMVY22 12k protein - potato virus M (strain Russian)
C:Species: potato virus M
A:Note: host Lycopersicon esculentum (tomato)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Jul-1999
C:Accession: P00003; C54333; S21603
R:Rupakov, V.V.; Morozov, S.Y.; Kanyuka, K.V.; Zavrilev, S.K.
J. Gen. Virol. 70, 1861-1869, 1989
A:Title: Partial nucleotide sequence of potato virus M RNA shows similarities
to potexviruses in gene arrangement and the encoded amino acid sequences.
A:Reference number: A92800; MUID:89293091
A:Accession: P00003
A:Molecule type: mRNA
A:Residues: 1-109 <RUP>
A:Cross-references: GB:D14449; GB:ID00515; GB:X53062; NID:9222424;
PIDN:BA03341.1; PID:dl003851; PID:922427
R:Zavrilev, S.K.; Kanyuka, K.V.; Levey, K.E.
J. Gen. Virol. 72, 9-14, 1991
A:Title: The genome organization of potato virus M RNA.
A:Reference number: A54333; MUID:91116326
A:Accession: C54333
A:Molecule type: genomic RNA
A:Residues: 1-109 <ZAV>
A:Cross-references: EXBL:X53062; NID:961291; PIDN:CA437334.1; PID:961294
A:Comment: The genome is a single-stranded, positive-sense RNA.
C:Superfamily: barley stripe mosaic virus 14k protein
C:Keywords: transmembrane protein
WMVY2 Length: 109 February 11, 2000 15:51 Type: P Check: 3415 ..

1 MPLTPPEPT KYVIALAGV SLALVWLLI RSTLPVVGDR DNNLPNGWY
51 RDGTRSVFYN SPGRINSTEA RKAPLIGPW AIYVLVLLI WASHKLGREN
101 CRACAGSHT

11AA_SEQUENCE 1.0
P1:VCBWC - coat protein - tobacco streak virus (strain WC)
N:Alternate names: P4 protein
C:Species: tobacco streak virus TSV
A:Note: host Nicotiana glauca X N. Clevelandii (tobacco)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 23-Jul-1999
C:Accession: A04206
R:Cornelissen, B.J.C.; Janssen, H.; Zuidema, D.; Bol, J.F.
Nucleic Acids Res. 12, 2427-2437, 1984
A:Title: Complete nucleotide sequence of tobacco streak virus RNA 3.
A:Reference number: A93507; MUID:84169544
A:Accession: A04206
A:Molecule type: genomic RNA
A:Residues: 1-237 <CON>
A:Cross-references: GB:X00435; NID:962146; PIDN:CA25133.1; PID:962148
C:Genetics:
A:Map position: segment 3
C:Superfamily: tobacco streak virus coat protein
VCBWC Length: 237 February 11, 2000 15:51 Type: P Check: 3595 ..

1 NNTLIGQPDH PSNAMSRRAN NRSNNNSCPT CIDEIDAMAR NCPAHNTVNT
51 VSRORRRANA RAAAYRRANA RVPLDLPVVS VSRPQAKSL RLPNNQWVT
101 RKASMSAKT VDTINDAIPFK TIVGIEPIEG AETKFRLLI GFVAVSDGE
```

```

151 GMDVGVTDV IDPPVYGR L GFKKTYRSR DFDLGSKLLN QLDRAVYWC
201 LDERRRRAKR VOLAGYIMAI SKPAPLMPE DELVND
!!AA_SEQUENCE 1.0
P1:VCVETC - coat protein - turnip crinkle virus
C:Species: turnip crinkle virus
A:Note: host Brassica rapa (turnip)
C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Jul-1999
C:Accession: JAO111; S07285
R:Carlington, J.C.; Heaton, L.A.; Zuidema, D.; Hillman, B.I.; Morris, T.J.
Virology 170, 219-226, 1989
A:Title: The genome structure of turnip crinkle virus.
A:Reference number: A94394; MUID:89233179
A:Accession: JAO111
A:Molecule type: genomic RNA
A:Residues: 1-351 <CAR1>
R:Carlington, J.C.; Morris, T.J.; Stockley, P.G.; Harrison, S.C.
J. Mol. Biol. 194, 265-276, 1987
A:Title: Structure and assembly of turnip crinkle virus. IV. Analysis of the
coat protein gene and implications of the subunit primary structure.
A:Reference number: S07285; MUID:87283926
A:Accession: S07285
A:Molecule type: genomic RNA
A:Residues: 1-351 <CAR2>
A:Cross-references: EMBL:X05193; NID:962114; PIDN:CAA28823.1; PID:962115
A:Note: the authors translated the codon TTG for residue 346 as Trp
C:Comment: The genome is a single-stranded, positive-sense RNA.
C:Superfamily: carnation mottle virus coat protein
C:Keywords: coat protein
F:1-56/Domain: R <DOR>
F:62-246/Domain: S <DOS>
F:249-351/Domain: P <DOP>
VCVETC Length: 351 February 11, 2000 15:51 Type: P Check: 1818 ..

1 MENDPRVAKF ASDGAQWAIK MOKKGMSTLT SROKOTARAA MGIRKSPVAQ
51 PQVKTRLSA PVLAAYREVS TOPRVSTARD GTRFSGSELI TLKKNITDTE
101 PRYTAVALNP SEPTGFNQLI KEAAQYEXYR FTSIRFRYSP MSPSTTGKRV
151 ALAPRDAAK PPPNDLASY NIEGCVSSVP WTGTLVLPV DSTDFFVADG
201 ISDPRLVDFG KLIMATYGOG ANDAQLGEV RVETVQLKN RFGSTSDAQI
251 GDFAGVKQCP RLYSMKTKG TAGMEHDCF LGTGNESLTL FTEKAPVSGL
301 ENADASDSV LGENAAGSVQ MAGVYAERG OGVMATTEE QKRGKQALR
351 I
!!AA_SEQUENCE 1.0
P1:VHPPF6 - major nucleocapsid protein - phage phi-6
N:Alternate names: P8 protein
C:Species: phage phi-6
A:Note: host Pseudomonas phage phi-6
C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999
C:Accession: A23368
R:McGraw, T.; Mindich, L.; Frangione, B.
J. Virol. 58, 142-151, 1986
A:Title: Nucleotide sequence of the small double-stranded RNA segment of
bacteriophage phi6: novel mechanism of natural translational control.
A:Reference number: A93017; MUID:86144085
A:Accession: A23368
A:Molecule type: genomic RNA
A:Residues: 1-149 <MCG>
A:Cross-references: GB:M12921; NID:9215492; PIDN:AAA32358.1; PID:9215493
A:Genetics:
A:Gene: P8
A:Map position: segment S
C:Superfamily: phage phi-6 nucleocapsid protein
C:Keywords: nucleocapsid

VHPPF6 Length: 149 February 11, 2000 15:51 Type: P Check: 6252 ..

1 MLFPVARRA VPAIESAIA TPGLYSRIA AIGSKYSPSA ILAAVKNPV
51 VAGLTLLAIG STGYDAYQOL LENHEVAEM LKDSLKADE IQDPFIGNLG
101 QYREELIVE DAARFVGMS NLIRLOALE LDIKIYGLKM QLNDMGTRS
!!AA_SEQUENCE 1.0
P1:ZGBP4 - gene G protein - phage G4
C:Species: phage G4
C>Date: 30-Nov-1979 #sequence_revision 30-Nov-1979 #text_change 23-Jul-1999
C:Accession: A04252
R:Godson, G.N.; Bartell, B.G.; Staden, R.; Fildes, J.C.
Nature 276, 236-247, 1978
A:Title: Nucleotide sequence of bacteriophage G4 DNA.
A:Reference number: A93200; MUID:79053264
A:Accession: A04252
A:Molecule type: DNA
A:Residues: 1-177 <GOD>
A:Cross-references: GB:J02454; GB:M10724; GB:M1404; GB:V00657; NID:9215415;
PIDN:AAA32324.1; PID:9215425
C:Comment: Gene G protein is one of the structural components of the
bacteriophage capsid.
C:Superfamily: phage phi-X174 gene G protein
ZGBP4 Length: 177 February 11, 2000 15:51 Type: P Check: 4484 ..

1 MQKRTSKHN APINSTOLAA TTPRAVAPV LSVPNLSRST ILINNTTAV
51 TTHSGICHHV RIDETNPTNH HALSTAGSL NVRADMIARA IFEVADGV
101 PTAVPALYDV YPIETFNNGK AISEKDAVTI DSHPRYVGD VYAGIMLSN
151 AMTASTISGV LSVQVNRRA TVLOPLK
!!AA_SEQUENCE 1.0
P1:PDYBC - REP2 protein - yeast (Saccharomyces cerevisiae) Plasmid
N:Alternate names: hypothetical protein C
C:Species: Saccharomyces cerevisiae
C>Date: 31-Oct-1980 #sequence_revision 23-Oct-1981 #text_change 23-Jul-1999
C:Accession: A04504
R:Hartley, J.L.; Donelson, J.E.
Nature 286, 860-864, 1980
A:Title: Nucleotide sequence of the yeast plasmid.
A:Reference number: A93232; MUID:81012161
A:Accession: A04504
A:Molecule type: DNA
A:Residues: 1-296 <HAR>
A:Cross-references: GB:J01347; NID:9172190; PIDN:AAB59340.1; PID:9172191
A:Experimental source: strain A364A D5
C:Genetics:
A:Genome: plasmid
C:Superfamily: yeast REP2 protein
C:Keywords: plasmid partition
PDYBC Length: 296 February 11, 2000 15:51 Type: P Check: 6596 ..

1 MDDIETAKNL TVKARIASV MDVCRIFIM IAPVDIDIE SKRKSDELLF
51 PGVIRPMES LTGRPPGLD SSAEDSVSS DSADEVILPA AKMKRRFES
101 IGNGLMSOE ASQAIDML QNNKLLDNK QLYSIAIII GRLEPKDKR
151 ATEMLRKMD CTOLLVPAP TEEDVAKLV VYTQLTLVP PDRQALLID
201 LFIPESLKDI FNSFNEAAE NRILOKSEL EGRETVNHN TNEVPSRRT
251 RSRDINARGA YKLQNTITEG PRAVPTKRR VATRGGRKS RNTSRV
!!AA_SEQUENCE 1.0
P1:B64166 - hypothetical protein H11084 - Haemophilus influenzae (strain Rd

```

KW20)
 C:Species: Haemophilus influenzae
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: B64166
 R:Feilichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
 Kellaway, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
 McEneaney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
 Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips,
 C.A.; Spillig, T.; Hedblom, E.; Cotton, M.D.; Usterbach, T.R.; Hanna, M.C.;
 Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
 Fuhmann, J.L.; Geoghagen, N.S.M.
 Science 269: 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
 Venter, J.C.
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.
 A:Reference number: A64000; MUID:95350630
 A:Accession: B64166
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-214 <TIGR>
 A:Cross-references: GB:U32788; GB:I42023; NID:g1574629; PID:g1574638;
 TIGR:HI1084
 C:Note: Best homolog was a hypothetical protein from Escherichia coli
 C:Superfamily: hypothetical protein HI1084
 B64166 Length: 214 February 11, 2000 15:51 Type: P Check: 1216 ..

1 MNIQLKWF TLTFVLTAFLVTRTAIAET SPYVLKQQA DRLFDIQAN
 51 OSKIKDOPNV LRTIVRNDLL PYVLKEXAGS KYLGSTYKST SAEQREKFEK
 101 TPEELIEQXV AQAALNYSNQ KIQEISEKEL GDNNFNINIV NIIQANGVAP
 151 ILIFYKRRKG NKSSEKMYVD MVGAGVSMLE DRIKNVGLL NKQIGDITLT
 201 KMQQASQPI IFNQ
 !:AA-SEQUENCE 1.0
 P1:E70044 - conserved hypothetical protein yvob - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: E70044
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,
 V.; Bartero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.;
 Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.;
 Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,
 J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzko, F.; Devine, K.M.;
 Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
 Fabret, C.; Ferrari, E.
 Nature 390: 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,
 G.; Galloway, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Goldschmidt, E.J.; Grand,
 G.; Gussuppi, C.; Guy, B.; Haga, K.; Halech, J.; Harwood, C.R.; Hentut, A.;
 Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;
 Joris, B.; Karamata, D.; Kasahara, Y.; Kjaer-Blanchard, M.; Klein, C.;
 Kobayashi, Y.; Koetter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;
 Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
 S.; Muesel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;
 Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,
 B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Portwolk, S.; Prescott,
 A.M.; Prescean, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds,
 S.; Rieger, M.; Rivolta, C.; Roeha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato,
 T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,
 A.; Seror, S.J.; Serro, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Tacconi, E.;
 Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka,
 T.; Terpiltra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenoel, M.;
 Vannier, F.; Vassarotti, A.; Viari, A.; Wambuti, R.; Medler, E.; Medler, H.;
 Weitenegger, T.; Wintner, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
 K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E70044
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-310 <KUN>
 A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PID:e1186188;
 PID:g2636013
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yvob
 C:Superfamily: hypothetical protein yvob
 E70044 Length: 310 February 11, 2000 15:51 Type: P Check: 4650 ..

1 MARVRKDVN EGFNLELSG EEGINRPTM SDLSRPGIEI AGFTTYPRP
 51 RVQLCKTEL SFEPQLPREE KQQRMDSLCT DTPAIIISR DMPIQDELID
 101 ASERKGVPLV RSPKTRRLS SRLTNFLESR LAPTAIHGV LVDYGVVL
 151 ITKSGVGKS ETALEVYRG HRLVADCV E IROEDQTLV GNAPELIEHL
 201 LERIGIGIIN VMTLFGAGAV RSNKRITIVM NLELMEOGQK YDRIGLEERT
 251 MKIIDIEIKR LTIIVRPGRN LAVIIEVAM NFRLRMGLN ABOFTKLA
 301 DVIDEGCEE
 !:AA-SEQUENCE 1.0
 P1:G69000 - molybdenum cofactor biosynthesis protein MoeA - Methanobacterium
 thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: G69000
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;
 Aldredge, T.; Bashirzaden, R.; Blakey, D.; Cook, R.; Gilbert, K.; Harrison,
 D.; Hoang, L.; Keagle, P.; Lumm, W.; Potchler, B.; Qiu, D.; Spadatoro, R.;
 Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.;
 Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.;
 Goyal, A.; Pietrovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.;
 Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta
 H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: G69000
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-409 <MTR>
 A:Cross-references: GB:AE000873; GB:AE000666; NID:g2622101; PID:g2622104
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1003
 A:Start codon: TTG
 C:Superfamily: molybdenum cofactor biosynthesis protein moeA-2
 G69000 Length: 409 February 11, 2000 15:51 Type: P Check: 7972 ..

1 MNGVEFLEIT DLDARRITG ELFSELSOG TESVDILDAD GRVLADIES
 51 PVDLPPFDRA SRDGYAVVAA DIFGADENP ATLRCIETVE AGVPSLEVA
 101 EGICTISTG APVEGADAV VAVEYTWEEY DRVYQRPAY PSOHIAARG
 151 DIRAGVLLR TGYLSPDKI ALSAAGITR VNVISKPSVC VISTGNELIE
 201 PSQQLPGRKI FDSNAGLSS ALKEACSVK HGGIYRDDYG ELRDALIRGL
 251 EEMDWITSG GTSAGAGIL SEVLDELGRV VIHGSIKPG KPIYGVVDD
 301 KIYVGLPGFP VSALIVYMSL IEPFIKLSG RKASGHVKT LKLAERLHSS

351 RGRHYALVM VGDYARPPIF KSGAVAYALA GADGITEVPR NFEIINEGE
401 VEVLFEENR

!!AA_SEQUENCE 1.0
PI:E70191 - conserved hypothetical protein BB0734 - Lyme disease spirochete
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: E70191
R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra,
R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty,
B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kesteven,
A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams,
M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Wathey, L.; McDonald, L.;
Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.;
Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genome sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
A:Reference number: A70100; PMID:96065943
A:Accession: E70191
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-337 <KLE>
A:Cross-references: GB:AE001173; GB:AE000783; NID:92688665; PID:92688669;
TIGR:BB0734
A:Experimental source: strain B31
C:Superfamily: *Bacillus subtilis* conserved hypothetical protein ywlc

E70191 Length: 337 February 11, 2000 15:51 Type: P Check: 8273 ..

1 MISTEISS OIOKAKLIK MGELVFEPE TVYGIGANAY NEDAVKIFL
51 VKRRPINPL IVHVDYVKKI KESEYIPKS ALMLIKRSP GPPLYVLKKS
101 IKSRVSGN LDTVAIRIPA NKTALSLIKA SKPIVAPSA NISKRPESTN
151 FEALKEUNG LVGIIKPEE NKDPNIGIES TVVGDLMDN VLLIRPGAIT
201 KKIENELOG KTVVYAEIK MELEKSPGNI IEHYKPKIPV YLRQSOINR
251 RYLNKOTKIL ITKATLKSYL FNFFNKKNI TYFNTLEEA QNLKKELYNS
301 ENNYKQISE FLKDEELGHS INNRIKKAAS NRIINKK

!!AA_SEQUENCE 1.0
PI:A71183 - probable UDP-glucose 4-epimerase - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A71183
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohnuki, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamaoka, J.; Kushioka, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.
A:Reference number: A71000; PMID:98344137
A:Accession: A71183
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-306 <KAW>
A:Cross-references: GB:AP000007; NID:93236134; PID:di0311799; PID:93258173
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by
Genbank
C:Genetics:
A:Gene: PH1742
C:Superfamily: hypothetical protein s111213

A71183 Length: 306 February 11, 2000 15:51 Type: P Check: 8100 ..

1 MKNKLIYIG GAGFIGSHA EALKDENDVI IIDNIXSRI ENIEPEGKFI
51 RADYDRESI AEVISEADIV FHEAQSIVK ESEIDPATE EVANVIGTIV
101 LRAISOQDGK LIFASSAAY GEKPELPIE DTLNIPISY GITKLAHEH
151 CRYQSILYGI PVIILRYENV YGRQSSAVA GVISIFLERA IKGEPLIIFG
201 DKGOTRPIYI VQDVVEANIL VAKKRSANGR IENVATGETI TIELANKII
251 DMTSSSSSIL FYPPRGDIR RSVAKIERIK KLGKPRRSL EELKETEFKX
301 FTSRTQ

!!AA_SEQUENCE 1.0
PI:S70955 - ctnf protein - *Vibrio cholerae*
C:Species: *Vibrio cholerae*
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S70955
R:Balk, E.M.; Bunschoten, A.E.; Willems, R.J.L.; Chang, A.C.Y.; Mool, F.R.
Mol Microbiol 20, 799-811, 1996
A:Title: Genetic organization and functional analysis of the ctn DNA essential
for cell wall polysaccharide synthesis in *Vibrio cholerae* O139.
A:Reference number: S70952; PMID:96386047
A:Accession: S70955
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-288 <BIR>
A:Cross-references: EMBL:X90547; NID:91462276; PID:el94952; PID:q1107920
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1995
C:Genetics:
A:Gene: ctnf
C:Superfamily: hypothetical protein b0985

S70955 Length: 288 February 11, 2000 15:51 Type: P Check: 8920 ..

1 MCOFDLFLRL RSPFKICRSP IKKRAWSFIT SSGWITLAAA ISGHSICVAA
51 PLATSLAVRV ANNOQAIHST RLSPSEPNRL DSAVQOTLEQ AGLTQOQVEW
101 PSAGLFLSLH AFLFKRDVLL KLAQOQSSAP PQOQALMAST IAOIROAEFA
151 KRLFISVDP WRIAPQHPN RLNSGWLTL NKSSTQVSIV GAVNQPDVVI
201 WHNLSAKDY ARAAGLIDQ ISELVIQPD GIAOKHAYV WNOFNEVAP
251 CAIVYVPLP KRAFPPIYV DADLNQVIE LERNRLPL

!!AA_SEQUENCE 1.0
PI:A70118 - glycine betaine, L-proline ABC transporter, permease protein (prow)
homolog - Lyme disease spirochete
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A70118
R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra,
R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwin, M.; Dougherty,
B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kesteven,
A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams,
M.D.; Gocayne, J.; Weidman, J.; Uterback, T.; Wathey, L.; McDonald, L.;
Artach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.;
Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genome sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
A:Reference number: A70100; PMID:96065943
A:Accession: A70118
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-299 <RLE>
A:Cross-references: GB:AE001125; GB:AE000783; NID:926886021; PID:926886023;

TIGR:BB0145
 A:Experimental source: strain B31
 C:Superfamily: glycine betaine/carnitine/choline ABC transporter
 A70118 Length: 299 February 11, 2000 15:51 Type: P Check: 6242 ..

1 MTKDFEIKTI DNEFDELVDN FSISDVGFS KSIIIEYENT KNLFLFVNPL
 51 LFIITVCLLS FVFLKKRLIF ILPGFEFII YFNLMXASMD TAIITFVSVL
 101 VSVILGIPIG ILGCEPFRFY VFLKPIIDLM QAMPPEFYLI PAIPFPGMGT
 151 ASAIFATIVF AMPPVIRYTR LGIVOVSDV IEAKSFGSS NDLIFOVOL
 201 PLISQSTIEG INOSIMMAIS MIVIAWVGS SGLGRVIYS IERLNGEGL
 251 ISGLAVYIA ILDRIMOSI FIKFSYLTND HYGKKNKRF KRFLFIYK

!!AA_SEQUENCE 1.0
 P1:S76375 - plastocyanin precursor - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein sl10139
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 11-Jun-1999
 C:Accession: S76375; S13733
 R:Kaneh, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosochi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimpo, S.; Tanuchi, C.; Wada, I.; Metanabe, A.; Yamada, M.; Tasuda, M.; Tabata, S.
 DNA Ref. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S76375
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-126 <EAN>
 A:Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BAA10227.1; PID:d1010878; PID:g1001599
 A:Experimental source: PCC 6803
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

R:Briggs, L.M.; Pecoraro, V.L.; McIntosh, L.
 Plant Mol. Biol. 15, 633-642, 1990
 A:Title: Copper-induced expression, cloning, and regulatory studies of the plastocyanin gene from the cyanobacterium Synechocystis sp. PCC 6803.
 A:Reference number: S13733; MUID:91338701
 A:Accession: S13733
 A:Molecule type: DNA
 A:Residues: 1-126 <BRI>
 A:Cross-references: EMBL:X54105; NID:g47401; PIDN:CAA38038.1; PID:g47402
 A:Experimental source: PCC 6803
 R:Romero, A.; De La Cerdá, B.; Varela, P.F.; Navarro, J.A.; Hervás, M.; De La Rosa, M.A.
 Submitted to the Brookhaven Protein Data Bank, June 1997
 A:Reference number: A68539; PDB:1PCS
 A:Contents: annotation; X-ray crystallography, 2.15 angstroms, 29-71, 'D', 73-76, 'P', 78-89, 'L', 91-126
 A:Note: engineered sequence expressed in Escherichia coli
 C:Genetics:
 A:Gene: petE
 C:Superfamily: plastocyanin
 C:Keywords: copper; electron transfer; metalloprotein
 F:1-18/Domain: signal sequence #status predicted <SID>
 F:19-28/Domain: propeptide #status predicted <PRO>
 F:29-126/Product: plastocyanin #status predicted <MAT>
 F:57,111,114,119/Binding site: copper (His, Cys, His, Met) (type 1) #status experimental

S76375 Length: 126 February 11, 2000 15:51 Type: P Check: 6628 ..

1 MSKFLITLIL GLLVVSFF LSVSPAAN ATYKMSDSG ALVEPSTVT

!!AA_SEQUENCE 1.0
 P1:F70646 - probable quinine oxidoreductase showing strong - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-Aug-1999
 C:Accession: F70646
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70646
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-323 <COL>
 A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PID:e291013; PID:g18126
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: fabB4
 C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
 F:23-310/Domain: long-chain alcohol dehydrogenase homology <LAOH>

F70646 Length: 323 February 11, 2000 15:51 Type: P Check: 8411 ..

1 MRAVVRTRLE GPDAVEAEV EEPISAGVVI EVNAGVAFP DALLTRGRYQ
 51 YRPEPFVGL AEIAGVYRSA PONSQVRSQD RVVGLTALWG GMAEVAVLSLP
 101 ERVEFLPDNM TFEAGAGVLF NDLFYFALA VRGRLQGET VLVHGAAGSI
 151 GSTLRLAPA LGASRTVAVY STOKERLAT VAGATDVLA EEFKDAVOEL
 201 TNGSGVDIVY DVGGBRFTD SLRSLAAGR LVIYGFGE IPTVKNRRL
 251 LNNIDVGVG WGAWSLTHPD ALAQWQSLE RLRSCKLPP PEPVYPIDQ

301 AAAIASLEN RTAKKRVLR VRD

!!AA_SEQUENCE 1.0
 P1:S01901 alcohol dehydrogenase (EC 1.1.1.1) 2 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 11-Jun-1999
 C:Accession: S01901
 R:Atkinson, P.W.; Mills, L.E.; Starmer, W.T.; Sullivan, D.T.
 Genetics 120, 713-723, 1988
 A:Title: Structure and evolution of the Adh genes of Drosophila melanogaster.
 A:Reference number: S01901; MUID:89137903
 A:Accession: S01901
 A:Molecule type: DNA
 A:Residues: 1-254 <ATK>
 A:Cross-references: EMBL:X12536; NID:g7553; PIDN:CAA31054.1; PID:g7554
 C:Genetics:
 A:Gene: Adh-2
 A:Cross-references: FlyBase:Fbgn0012567
 A:Introns: 31/3; 166/3
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: alcohol metabolism; NAD; oxidoreductase
 F:6-182/Domain: short-chain alcohol dehydrogenase homology <SADH>

S01901 Length: 254 February 11, 2000 15:51 Type: P Check: 8887

1 MAIANKNILF VAGLGIGFD TSREIVKSGP KNVLILDRIE NPAIAELKA

51 LNFRTVTFE PYDVTVSVAE TTKLKTIFD KIKTVLDLIN GTGLDDHOI

101 ERTIAVNETG TLNNTTAIMS FMDKRCGPG GTIANICSVT GFNNILVVP

151 YSASKAAALS FTNSLARLAP ITGVATYSIN PGITRTTLVH KFNGLVDEP

201 RVALLLEHP TOTLQCAQN FVAKIQANON GAIWMLDGT LEALEWTKKH

251 DSHI

11AA_SEQUENCE 1.0

F1:S65732 - carbonyl reductase (NADPH) (EC 1.1.1.184) - chicken (fragments)

C:Species: Gallus gallus (chicken)

C>Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998

C:Accession: S65732

R:Tsai, S.P.; Wang, L.Y.; Yeh, H.I.; Tam, M.F.

Biochim. Biophys. Acta 1292, 288-292, 1996

A:Title: The 30 kDa protein co-purified with chick liver glutathione S-transferases is a carbonyl reductase.

A:Reference number: S65732; MUID:96176857

A:Accession: S65732

A:Status: Preliminary

A:Molecule type: Protein

A:Residues: 1-13;14-42;43-58;59-105;106-127;128-132 <TSA>

C:Superfamily: Ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: Oxidoreductase

F:4-35/Domain: short-chain alcohol dehydrogenase homology (fragments) <SADH>

S65732 Length: 132 February 11, 2000 15:51 Type: P Check: 4453

1 SNVPAVVTG TNKPXGQEA VAKLOEGLH PLFHOOLDID LQVSDTPPA

51 VOAEVTLKGV HEKEGMPNSA YGVSKIGTV LSRIOARMAN ESKGPHILL

101 NAXXPAPKSP EGAETPVYL ALPPDAIVR TW

11AA_SEQUENCE 1.0

P1:JC284 - carbonyl reductase (NADPH) (EC 1.1.1.184), Inducible - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 16-Apr-1997 #sequence_revision 18-Jul-1997 #text_change 11-Jun-1999

C:Accession: S68983; PC2234; JC284; PC2159; S52349

R:Wernuth, B.; Maeder-Heinmann, G.; Ernst, E.

Eur. J. Biochem. 228, 473-479, 1995

A:Title: Cloning and expression of carbonyl reductase from rat testis.

A:Reference number: S68983; MUID:95220378

A:Accession: S68983

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-277 <WER>

A:Cross-References: EMBL:X8349; NID:9666086; PIDN:CA53088.1; PID:9666087

R:Rolt, E.; Soederstrom, M.; Ahlberg, M.B.; DePierre, J.W.

Biochem. Biophys. Res. Commun. 201, 149-154, 1994

A:Title: A novel 34kDa glutathione-binding protein in mature rat ovary.

A:Reference number: PC2159; MUID:94256971

A:Accession: PC2234

A:Molecule type: Protein

A:Residues: 104-121,'X',123-134,'D',136-137,'H' <TOS>

A:Experimental source: ovary

R:Okil, H.; Okada, T.; Mizutani, T.; Numata, Y.; Mingsishi, T.; Miyamoto, K.

Biochem. Biophys. Res. Commun. 230, 518-523, 1997

A:Title: Identification of two closely related genes, inducible and noninducible carbonyl reductases in the rat ovary.

A:Reference number: JC5284; MUID:97167755

A:Contents: ovary

A:Accession: JC5284

A:Molecule type: mRNA

A:Residues: 1-140,'GW',143,'R',145-235,'T',237-238,'E',240-277 <OK>

A:Cross-References: DDBJ:D89069; NID:9106811; PIDN:BA19007.1; PID:dl019750; PID:g1906812

C:Comment: This enzyme is a cytosolic, monomeric oxidoreductase that catalyzes the NADPH-dependent reduction of variety of endogenous carbonyl compounds.

C:Genetics:

A:Gene: 1CR

C:Superfamily: Ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: cytosol; monomer; NADP; oxidoreductase

F:6-185/Domain: short-chain alcohol dehydrogenase homology <SADH>

JC5284 Length: 277 February 11, 2000 15:51 Type: P Check: 7185

1 MSSDRPALV TGANKSIGFA IVRDLCKREL GDVVLARDE SRGHEAVQL

51 QTEGLSPRFH QLDIDNPQI RALRDLFLE YGGLNLYVN AGIAFKVVD

101 TPFHQAQVT MKNTEFGTD VCKELLPILK PQGRVNVSS SVSLRLKSC

151 SPELOQKERS ETITEELVG LMKFIEDPAK KGVHAKGWP NSAVGVTKIG

201 VTLSRIYAR KLINEEREDK ILLNACPGW VRTDMAGPKA TKSPEGAE

251 FVYIALPFG AEGPHQGFVQ DKVEPW

11AA_SEQUENCE 1.0

P1:S54815 - undamycin polyketide ketoreductase (EC 1.1.1.-) - Streptomyces fradiae

C:Species: Streptomyces fradiae

C>Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 11-Jun-1999

C:Accession: S54815

R:Decker, H.; Haag, S.

submitted to the EMBL Data Library, May 1995

A:Description: Cloning and characterization of a polyketide synthase from Streptomyces fradiae 7ue2717 encoding the genes for biosynthesis of the undamycin antibiotic undamycin A and a gene probably involved in its oxygenation

A:Reference number: S54810

A:Accession: S54815

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-261 <DEC>

A:Cross-References: EMBL:X87093; NID:9809102; PIDN:CA60572.1; PID:9809108

C:Superfamily: Ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: NAD; oxidoreductase

F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>

S54815 Length: 261 February 11, 2000 15:51 Type: P Check: 5190

1 MTOONPRVAL VTGATSGIGL AVTRILGRQG HAVFLCART DSVSTVKOL

51 LDEGLEVDGA PCDVRSADVD ERFVORAVHR FETIDVLVNN AGRGGGVT

101 DIADELWHDV IDTNLSVFR MTRVINTGS MRHKDRGRII NIASTAKOG

151 VVLGAPYSAS KHGVVGFKA LGNELAPPTGI TYNAVCPQIV ETPMAQVRA

201 GYAAADTSE DAILEKPAK IPIGRSTPE EVAGLVGILA SDTNAASITSO

251 ALNVCGLGN F

11AA_SEQUENCE 1.0

P1:D5587 - griseusin polyketide ketoreductase (EC 1.1.1.-) - Streptomyces griseus

C:Species: Streptomyces griseus

C>Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 11-Jun-1999

C:Accession: D5587

R:Yu, T.W.; Bibb, M.J.; Revill, W.P.; Hopwood, D.A.

J. Bacteriol. 176, 2627-2634, 1994

A:Title: Cloning, sequencing, and analysis of the griseusin polyketide synthase gene cluster from Streptomyces griseus.

A:Reference number: A55587; MUID:94222841

A:Accession: D5587

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-261 <YUA>

A:Cross-references: GB:J77865; NID:6488770; PIDN:CAA54861.1; PID:6488774
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>
D5587 Length: 261 February 11, 2000 15:51 Type: P Check: 3920 ..

1 MSDODRVAV VTGATSGIGL SVARLLASOG HRVFIGARSA ENVAATKEL
51 VAEGLDVDA VDVRSDADV EAWIOAAVDR EGTVDVYVNN AGRSGGVTA
101 DIDPELMHDV IDTNLSVFR VTRSAINTGG LRSKSRGRII NIATSGKQG
151 VILGAPYSAS KHGVGFTKA LGNELAPTGI TVNAVCPGYV ETPMAQRVQ
201 GYAAAYDATE DAILEKFOAK IPLGRYSTPD EVAGLVGYLA SDTAASITSQ
251 ALNVCGGLGN F

!!AA_SEQUENCE 1.0
P1:S69225 - nogalamycin polyketide ketoreductase (EC 1.1.1.-) snod -
C:Streptomyces nogalater
C:Species: Streptomyces nogalater
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 17-Mar-1999
C:Accession: S69225
R:Ylthorke, K.; Tuikka, J.; Jussila, S.; Cong, L.; Meentsaelae, P.
Mol. Gen. Genet. 251, 113-120, 1996
A>Title: A gene cluster involved in nogalamycin biosynthesis from Streptomyces
nogatater: sequence analysis and complementation of early-block mutations in
the anthracycline pathway.
A:Reference number: S69223; MUID:96242142
A:Accession: S69225
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <YLI>
A:Cross-references: EMBL:Z48262; NID:q1181156; PID:e221092; PID:q1181158
A:Experimental source: ATCC 27451
A>Note: the nucleotide sequence was submitted to the EMBL Data Library,
February 1995
C:Genetics:
A:Gene: snod
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:7-189/Domain: short-chain alcohol dehydrogenase homology <SADH>
S69225 Length: 262 February 11, 2000 15:51 Type: P Check: 7076 ..

1 MPTDGRVAV VTGATSGIGL AVYKALATG YACRSICARS QDAVDSTVEE
51 LRDEKLEVDG RSCDVSTDD VREFTGATR RFGGVYLVN NGRNGGGPT
101 ARITDELWLD VIETNLTSVF RVTREVLTTG GMLAAGRRI INIATSGKQ
151 GVILGAPYSA SKHGVGFTK ALGLELAKSG ITVNAVCPGY VETPAGERVG
201 RGYAEVMTS EEAVERFOA KIPLEGYVAP EYVAGLVTYL TPTADISITA
251 QALNVCGGLGN NY

!!AA_SEQUENCE 1.0
P1:S35196 - hypothetical protein 3 - Saccharopolyspora hirsuta
C:Species: Saccharopolyspora hirsuta
C>Date: 10-Dec-1993 #sequence_revision 21-Jul-1995 #text_change 28-May-1999
C:Accession: S35196
R:Le Gouill, C.; Desmarais, D.; Dery, C.V.
Mol. Gen. Genet. 240, 146-150, 1993
A>Title: Saccharopolyspora hirsuta 367 encodes clustered genes similar to
ketocycl synthase, ketocycl reductase, acyl carrier protein, and biotin
carboxyl carrier protein.
A:Reference number: S35194; MUID:93341453
A:Accession: S35196
A>Status: Preliminary
A:Molecule type: DNA

A:Residues: 1-261 <GON>
A:Cross-references: GB:M88258; NID:9347176; PIDN:AAA26487.1; PID:9347179
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:7-188/Domain: short-chain alcohol dehydrogenase homology <SADH>
S35196 Length: 261 February 11, 2000 15:51 Type: P Check: 9329 ..

1 MAANTRPVAL VTGATSGIGL AIRRLGQUG ARVYICARRA EELATTIKTL
51 QAEGLVDSG TDVSVPEQI PELVAAVDR EGPIDVLVNN AGRSGGVTV
101 EYDETWEDV INTNLNVFL MTKOVLTKG MLEDERGRII NIATSGKQG
151 VINGSPYSAS KHGVGFTKA LGLELAKTGI TVNAVCPGYV ETPMAQRVQ
201 HYSKIVSVE AEADRVTR VPLGRVPEE EVAAVEYLV SDGAAAVTAQ
251 ALNVCGGLGN Y

!!AA_SEQUENCE 1.0
P1:D64051 - 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) -
Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 11-Jun-1999
C:Accession: D64051
R:Fielschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips,
C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;
Nuyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
Fuhman, J.L.; Geophagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
Venter, J.C.
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: D64051
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-242 <TIGR>
A:Cross-references: GB:U22701; GB:L42023; NID:q1573105; PIDN:AAC21824.1;
PID:q1573112; TIGR:H10135
C:Genetics:
A:Gene: tabg
C:Function:
A:Pathway: fatty acid biosynthesis
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
C:Keywords: fatty acid biosynthesis; NADP; oxidoreductase
F:4-180/Domain: short-chain alcohol dehydrogenase homology <SADH>
F:4-34/Region: beta-alpha-beta NADP nucleotide-binding fold
F:149/Active site: Tyr #status predicted
D64051 Length: 242 February 11, 2000 15:51 Type: P Check: 4159 ..

1 MOKIALVVG STGIGRAIA EELSKGAFV IGTATSEKA EATSAVLGDK
51 GKGLVAVTD KESIEITLEDQ IKNDFGIDI LVNNAGITRD NLIMRKDEE
101 WEFIMQNTLT SVYHLSKRL RSMKRRFRG IINIGSVGS TGNPGQTNVC
151 AAKAGVGS KSLAKEVAR GITVNVAVG FIANDMEVL TDQRKGILS
201 NVPAGRLGEA KDIKAVAFI ASDDAVYIIG TLLHNGGLY LS

!!AA_SEQUENCE 1.0
P1:D66648 - 2-keto-3-deoxygluconate oxidoreductase kdud - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
C:Accession: D66648
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,
V.; Bertero, M.G.; Besterles, F.; Boiotin, A.; Borchert, S.; Boriss, R.;

Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, R.M.; J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denzot, F.; Devine, K.M.; Duesterhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Errington, J.; Fabeet, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galleon, N.; Gilm, S.V.; Glaser, P.; Goffeau, A.; Goldlight, E.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Klein, C.; Joris, B.; Karamata, D.; Kashara, Y.; Klserr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogg, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, A.; Authors: Lauber, U.; Lazarevic, V.; Lee, S.M.; Moll, R.P.; Mizuno, M.; Moestl, D.; S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Muzum, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Potwolk, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rlyolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sotokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamashita, M.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, M.; Vannier, F.; Vassaret, A.; Varti, A.; Wandt, R.; Wedler, E.; Wedler, H.; Wittenberger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69648
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1254 <UNT>
 A:Cross-references: GB:Z99115; GB:AL009126; NID:92634478; PID:CA81431.1; PID:el18360; PID:82634633
 A:Experimental source: strain 168
 A:Genetics:
 A:Gene: kdsL
 A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 F:13-190/Domain: short-chain alcohol dehydrogenase homology <SADH>

D69648 Length: 254 February 11, 2000 15:51 Type: P Check: 3484 ..

1 MGYLHDAFSL KGTALVTGP GTGIGGCIAT ALAGAGADI GTSHTSLSE
 51 TQOLVEDGR IFTSTLDS KPAIKDSAA ELFENQIDI LVNAGIHR
 101 EKADPEEN WOHVNVNIN SLFITOLAG RMLKRGHR IINIASLSF
 151 OGGLVPAYT ASKHAVAGIL KSFANEMAS GIOVNAIADG YISTRANKPI
 201 RDEKRNEDI LKRIPAGRW QADDIGTAV FLASRASDYV NGHILAVDG
 251 WLSR

!!AA_SEQUENCE 1.0
 P1:CA8674 - triophone reductase homolog - jmsoneed
 C:Species: *Datura stramonium* (jmsoneed, common thornapple)
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 11-Jun-1999
 C:Accession: CA8674
 R:Nakajima, K.; Hashimoto, T.; Yamada, Y.
 Proc. Natl. Acad. Sci. U.S.A. 90, 9591-9595, 1993
 A:Title: Two triophone reductases with different stereospecificities are short-chain dehydrogenases evolved from a common ancestor.
 A:Reference number: A8674; MUID:94022421
 A:Accession: CA8674
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-268 <NAK>
 A:Cross-references: GB:U20475; NID:9424157; PID:AAA33280.1; PID:9424158
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

F:18-198/Domain: short-chain alcohol dehydrogenase homology <SADH>

CA8674 Length: 268 February 11, 2000 15:51 Type: P Check: 5293 ..

1 MAGREIGGCG RRMSLRGMTA LVYGTGRGIG YAIYEELANF GAEVYTCSRS
 51 QNDDELEK WRKQFVSG PVCVSSISO ROTLMEVTS SENGKINILI
 101 NNAGTIPEK ATNFTAEDS IINGTFEAS YNLCLAHPL LKASGNASIV
 151 FNSAAAVIA VLSSIVAS KGAINGVTS LACEWAKDSI RVNAVAPWII
 201 NPIPEACQ VPSQKNIES LIGRAPKRA GPSEVSSLV TYICLPTASY
 251 ITGOTICVDG GTYVNGFI

!!AA_SEQUENCE 1.0
 P1:112020 - 3-oxoacyl-acyl carrier protein reductase - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
 C:Accession: 112020
 R:Kutchna, A.J.; Hoang, T.T.; Schweizer, H.P.
 Submitted to the EMBL Data Library, February 1997
 A:Description: Molecular characterization of the acpP-containing region of the *Pseudomonas aeruginosa* chromosome.
 A:Reference number: 217380
 A:Accession: 112020
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-247 <KUT>
 A:Cross-references: EMBL:U91631; NID:92738151; PID:92738155
 C:Genetics:
 A:Gene: fabG
 A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

112020 Length: 247 February 11, 2000 15:51 Type: P Check: 9922 ..

1 MSIGKVALY TGASRGIGQA IALELGRUGA VVIGTATSAS GAETIAETLK
 51 ANGVAGAGLV LDVSSDESVA ATLEHQHQL GQPLIVVNA GITRDNELVR
 101 MKDDEFDVY NTNINSLYRL SKAVLRGMK ARWGRINIG SVVAGMGNAG
 151 QTVAAKAG LEGTRALAR EVGSRATVY AVARGFIDTD MRELPEDAR
 201 EALIGQIPUG RLGQAEETAK VVGFLASDGA AVVIGATVPV NGMWMS

!!AA_SEQUENCE 1.0
 P1:112051 - 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) - *Vibrio*
 C:Species: *Vibrio* harveyi
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
 C:Accession: 112051
 R:Shen, Z.; Byers, D.M.
 J. Bacteriol. 178, 571-573, 1996
 A:Title: Isolation of *Vibrio* harveyi acyl carrier protein and the fabG, acpP and fabF genes involved in fatty acid biosynthesis.
 A:Reference number: 217396; MUID:96134997
 A:Accession: 112051
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-244 <SHD>
 A:Cross-references: EMBL:U39441; NID:91173839; PID:91173841
 A:Experimental source: strain B392
 A:Genetics:
 A:Gene: fabG
 A:Function: fatty acid biosynthesis
 A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: fatty acid biosynthesis; MDP: oxidoreductase

112051 Length: 244 February 11, 2000 15:51 Type: P Check: 5984 ..

1 MNEKIALV TGASRGIGRA IAEVLVERGA TVIGRTASEG GAALISEYLG
 51 ENGKIALNV TDVESIEATL KTIINDECGAI DILVNNACIT RDNLIMRKMD
 101 DEMNDIINTN LTPYIRMSKA VLKGMKKRRA GRINVGSSV GTMGNAQOTN
 151 YAAKAGVIG FTKSMAREVA SRGVTYNVA POFIETDMTK ALNDQDRAAT
 201 LSNVPAGRLG DPREIASAVV FLASPPAAVY TGETLHVNGG MYMV

!!AA.SEEQUENCE 1.0
 P1:S39654 - hypothetical protein X - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S39654
 R:Krim, M.; Bally, M.; Ball, G.; Tommassen, J.; Teerink, H.; Filloux, A.; Lazdunski, A.
 MO: Microbiol. 10, 431-443, 1993
 A>Title: Acp-mediated protein secretion in Pseudomonas aeruginosa: identification of two additional genes and evidence for regulation of xcp gene expression.
 A:Reference number: S39652; MUID:95020542
 A:Accession: S39654
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-255 <AKR>
 A:Cross-references: EMBL:X68594; NID:9431183; PID:9581439
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology F:11-187/Domain: short-chain alcohol dehydrogenase homology <SADH>
 S39654 Length: 255 February 11, 2000 15:51 Type: P Check: 5853 ..

1 MIEPINKNG KVALYTGAR GIGLISANL IAEQOVVLA DNDREKARY
 51 AEALEGHAF VAMDAQEGQ VANSVAEVLG QFGRIDGLVC NAALINRNT
 101 PLEALSIGEM TRTLAVNTLG PMLAYCTP YIRAHNGAIV NIASTRHOS
 151 EPDSAEYAAK KGLIALTHA LAASLPDIR VNALSFGWID TREAREEAA
 201 PLEILDHDOH LVGRVYED VASLVAMLS EDAGFVTQGE FLYDGMTRK
 251 MYILD

!!AA.SEEQUENCE 1.0
 P1:H69885 - 3-oxoacyl- acyl-carrier protein reductase homolog ymfI - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 C:Accession: H69885
 R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.; Bolochin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Broiliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleon, N.; Ghm, S.Y.; Glaeser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Guiller, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hult, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Priesen, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solio, B.; Sorokin, A.; Taconi, E.; Takagi, T.; Takahashi, H.; Takekura, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vanlier, F.; Vassartelli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: H69885
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-242 <KUN>
 A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PID:e1185278; PID:92634059
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ymfI
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology F:3-181/Domain: short-chain alcohol dehydrogenase homology <SADH>
 H69885 Length: 242 February 11, 2000 15:51 Type: P Check: 6790 ..

1 MNKALITGA SCGIGKISE TLAARGVNL LHYNTQNNAA AELAEKLSOM
 51 FGVAETIQA DLSAQDADR LTSSIVPID AIVLNGRSH FGLITVDNA
 101 TVQENVQOHV ASPYMLTRNL LPMGRNKSQ AIYAVSSING EFGASCEVLY
 151 SMAGQACQSF VKLAKELAP SGIRNNAVAP GAVDTMMNQ FTPEAKKEIA
 201 DEIPGRGR TORIADATAP LISEKASYIT GQLISNGCH HC

!!AA.SEEQUENCE 1.0
 P1:H69868 - glucose 1-dehydrogenase homolog ykwo - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
 C:Accession: F69868
 R:Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bertero, M.G.; Bessières, P.; Bolochin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Broiliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleon, N.; Ghm, S.Y.; Glaeser, P.; Goffeau, A.; Golightly, E.J.; Grandt, G.; Guiller, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hult, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Priesen, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleicher, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Solio, B.; Sorokin, A.; Taconi, E.; Takagi, T.; Takahashi, H.; Takekura, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vanlier, F.; Vassartelli, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

Mon Feb 14 08:07:32 2000

plr.cat

Page 71

A:Reference number: A69580; MUID:98044033
A:Accession: F69868
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-248 <RDN>
A:Cross-references: GB:499111; GB:AL009126; NID:g2633699; PID:el184967;
PID:g2633748
A:Experimental source: strain 168
C:Genetics:
A:Gene: ykvo
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F:7-182/Domain: short-chain alcohol dehydrogenase homology <SADH>
F69868 Length: 248 February 11, 2000 15:51 Type: P Check: 7200 ..

1 MGKEGKIAL VTGTSIGL ATAOKFVNEG AYYITGRQ NELDAVNCI
51 GKNYGVGD ISKLELDKL YDIKQKQK LDILFANAGI GNFPLGEIT
101 EEOVDRETDI NVKGIETVQ KALSLEPDKV GSIIITGSTA GSIGNPASFV
151 YGASKAALRA LVKNWILDK GTEIRNVVS PGSLTPAYD ELFGDALBEV
201 LENSNTVPA GKVGTEEEVA NAVSPLASDE SSYLTVLEL VDGILACV

11AA_SEQUENCE 1.0
P1:B55850 - hypothetical protein (hsdh 5' region) - Clostridium sorde111
C:Species: Clostridium sorde111
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 26-Aug-1999
C:Accession: B55850
R:Colman, J.P.; Hudson, L.L.; Adams, M.J.
J. Bacteriol. 176, 4865-4874, 1994
A:Title: Characterization and regulation of the NADP-linked
7alpha-hydroxyesteroid dehydrogenase gene from Clostridium sorde111.
A:Reference number: A55850; MUID:94373451
A:Accession: B55850
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <COL>
A:Cross-references: GB:112058; NID:g431280; PID:g431281
A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
F:7-189/Domain: short-chain alcohol dehydrogenase homology <SADH>
B55850 Length: 267 February 11, 2000 15:51 Type: P Check: 1047 ..

1 MNKLEKVAL VTSAPRGICL ASAIRKLAONG AIYMGVRL EATQETCDKY
51 KEELILAKPV FPDVAINIDIT KEMIDTIRK ESKIDILVNN PETGPEKDL
101 DLVNGDEDTF PELFNVNGS VYRLSKLIIP HMIENKGGI VNISVGGSI
151 PDIRIRIGV SKSGVNNIRK QIAIOYAKG IRCNAVLPGL IATDAANSM
201 PDERRKFLS HYLNRICNP EDIANSVLEF VPESDSYIT GSILEVSGCY
251 NLGTPQIAEF VSGSKYE

11AA_SEQUENCE 1.0
P1:A56275 - 1,3-propanediol dehydrogenase (EC 1.1.1.202) - Citrobacter freundii
C:Species: Citrobacter freundii
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 11-Jun-1999
R:Daniel, R.; Bogenigk, R.; Gottschalk, G.
J. Bacteriol. 177, 3151-3156, 1995
A:Title: Purification of 1,3-propanediol dehydrogenase from Citrobacter
freundii and cloning, sequencing, and overexpression of the corresponding gene
in Escherichia coli.
A:Reference number: A56275; MUID:95238288
A:Accession: A56275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-387 <DAN>
A:Cross-references: GB:U09771; NID:g1229153; PIDN:AB4848.1; PID:g493086

C:Genetics:
A:Gene: dhdt
C:Superfamily: lactaldehyde reductase; lactaldehyde reductase homology
C:Keywords: homocitramine; oxidoreductase
F:8-383/Domain: lactaldehyde reductase homology <LAR>
A56275 Length: 387 February 11, 2000 15:51 Type: P Check: 101 ..

1 MSYRMDIYLV PRVNFEPNA ISVGERCKL LGKRALVLT DKGLRAIKDS
51 AVDKITLHNR EAGIDVYVFD GVEPNKRDN VADGLEVERK EHCIDIYVAG
101 GGSPPDCGKG IGIAATHEGD LYSVAGIELL TNPUPPIYAV NTTAGTASEV
151 TRHCVLNTR TKVRFVYSW RNLPSVIND PLMLGKRFAP LTAATGMDAL
201 THAVEYISK DANPTDAAA IQAIRLIARN LRQAVALGSN LKARENNAYA
251 SLAGMAFNN ANLGIVHAMA HÖLGGLYDMP HGVAVALLP HVARYNLIAN
301 PEKADIAEF MGENTDGLST MDAEELAIHA IARLSADIGI PÖHLRDLGVK
351 EADFPYMAEM ALKQDNAFSN PRKGNKEITA EIFROAF

11AA_SEQUENCE 1.0
P1:S47829 - glycerol-3-phosphate dehydrogenase (NAD+) (EC 1.1.1.8) -
Escherichia coli
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 11-Jun-1999
C:Accession: S47829; B65161
R:Plunkett, G.
Submitted to the EMBL Data Library, March 1994
A:Reference number: S47829
A:Accession: S47829
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-339 <PLD>
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AB18585.1; PID:g1657692
R:Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Colliado-Vides, J.; Glaesner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B65161
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <BLAT>
A:Cross-references: GB:AB000439; GB:U00096; NID:g1790036; PIDN:AAC76632.1;
PID:g1790037; UMG: B3608
A:Experimental source: strain K-12, substrain MG1655
A:Gene: gpsA
C:Genetics:
C:Superfamily: glycerol-3-phosphate dehydrogenase
C:Keywords: oxidoreductase
S47829 Length: 339 February 11, 2000 15:51 Type: P Check: 4443 ..

1 MNQRNASMTV IGAGSYGTAL AITLARNGHE VLMGHDPHE IATLERDRN
51 AAFLEDPVPP DTLHLESDLA TALAASRNIL VVBSHVGE VLKQIKPLMR
101 PPARLVMAWK GLEAETGRLL ODVAREALGD QIPLAVISGP TFARKEIACL
151 PRAISLASTD QTEADDIQOL LHCGKSFERY SNPDFIGVOL GGAVKNVIAI
201 GAGMSDGIKF GANARALIT RGLAEMSRIG AALGADPAIF MGWAGIGDLV
251 LCTDNQSRN RREGMIGOG MDVQSAQEKI GÖVVEGYRNT KEVELAHRE
301 GVEMPTTEI YÖVLYCKNA REAALTLGR ARKDESSSH
301 GVEMPTTEI YÖVLYCKNA REAALTLGR ARKDESSSH

11AA_SEQUENCE 1.0
P1:A44132 - D-aspartate oxidase (EC 1.4.3.1) - bovine

C:Species: Bos primigenius taurus (cattle)
 C:Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 21-Aug-1998
 A:Accession: A44132
 R:Negri, A.; Cecciani, F.; Tedeschi, G.; Simonic, T.; Ronchi, S.
 J. Biol. Chem. 267, 11865-11871, 1992
 A:Title: The primary structure of the flavoprotein D-aspartate oxidase from beef kidney.
 A:Reference number: A44132; MUID:92291057
 A:Accession: A44132
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-338 <NEG>
 A:Experimental source: kidney
 A:Note: sequence extracted from NCBI backbone (NCBIP:106605)
 C:Superfamily: D-amino-acid oxidase
 C:Keywords: oxidoreductase

A44132 Length: 338 February 11, 2000 15:51 Type: P Check: 6742 ..

1 MDTVRIAVVG AGVGLSTAV CISKVPGCS ITVIDKFTF ETTSDVAAGM
 51 LIPIPTDPTP IQKQKWFKE TPDHLEAIIV SAEADAGVI LVSGWQIFOS
 101 IPPEEVPYVA DVLGFRKMT KDEKKFPQH VFGHAFITLK CEGPVIPLW
 151 QKRVKNGGL ILTRIEDLM ELHPSFDIV NCSGLSROL AGDSKIFPV
 201 GOVLKQVAPV VKHFIRDSG LTIYIPGVN VILGTRQKG DMLSPDAEI
 251 SKETLSRCA LEPSLRGAYD LREVRGLRPI RGVLELEKL LAODSRRLPV
 301 VHHYHGSGG IAHMWGTALF ATRLVNCEVO VLRTPAPK
 !!AA_SEQUENCE 1.0
 P1:JC5438 - D-aspartate oxidase (EC 1.4.3.1) type III 1 - human
 N:Alternate names: aspartic oxidase
 C:Species: Homo sapiens (man)
 C:Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 11-Jun-1999
 C:Accession: JC5438
 R:Setoyama, C.; Miura, R.
 J. Biochem. 121, 798-803, 1997
 A:Title: Structural and functional characterization of the human brain D-aspartate oxidase.
 A:Reference number: JC5438; MUID:97306065
 A:Accession: JC5438
 A:Molecule type: mRNA
 A:Residues: 1-341 <SET>
 A:Cross-references: DDBJ:D89858; NID:91742023; PIDN:BAA14031.1; PID:d1014731; PID:q1742024
 A:Experimental source: brain
 A:Note: the authors translated the codon GAA for residue 83 as Asp and TGG for residue 95 as Leu
 C:Comment: This enzyme is a flavoprotein that catalyzes the oxidative deamination of dicarboxylic D-amino acids to give the corresponding alpha-keto acids and ammonia. It is involved in the regulation of the central nervous system and/or the development of the brain.
 C:Superfamily: D-amino-acid oxidase
 C:Keywords: oxidoreductase
 F:223/Active site: Tyr #status predicted
 F:278/Active site: Arg #status predicted

JC5438 Length: 341 February 11, 2000 15:51 Type: P Check: 7167 ..

1 MDTVRIAVVG AGVGLSTAV CISKVPGCS ITVIDKFTF ETTSDVAAGM
 51 LIPIPTDPTP IHTOKWFRE TFNHLFAIAN SAEADAGVH LVSGWQIFOS
 101 TPPEEVPYVA DVLGFRKMT EAEKKFPQY VFGQAFITLK CEGPVIPLW
 151 EKIKSGGM ILTRIEDLM ELHPSFDIV NCSGLSROL AGDSKIFPV
 201 GOVLKQVAPV VKHFIRDSG LTIYIPGVN VILGTRQKG DMLSPDAEI

251 SREILSRCA LEPSLHGACN IREKVGRLPY RGVRLQTEL LARDGRLPV
 301 VHHYHGSGG ISVHWGTALF AARLVSECV ALRTPIPKSN L
 !!AA_SEQUENCE 1.0
 P1:JC5439 - D-aspartate oxidase (EC 1.4.3.1) type III 2 - human
 N:Alternate names: aspartic oxidase
 C:Species: Homo sapiens (man)
 C:Date: 17-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 07-May-1999
 C:Accession: JC5439
 R:Setoyama, C.; Miura, R.
 J. Biochem. 121, 798-803, 1997
 A:Title: Structural and functional characterization of the human brain D-aspartate oxidase.
 A:Reference number: JC5438; MUID:97306065
 A:Accession: JC5439
 A:Molecule type: mRNA
 A:Residues: 1-282 <SET>
 A:Cross-references: DDBJ:D89858
 A:Experimental source: brain
 A:Note: the authors translated the codon GAA for residue 83 as Asp
 C:Comment: This enzyme is a flavoprotein that catalyzes the oxidative deamination of dicarboxylic D-amino acids to give the corresponding alpha-keto acids and ammonia. It is involved in the regulation of the central nervous system and/or the development of the brain.
 C:Superfamily: D-amino-acid oxidase
 C:Keywords: oxidoreductase

JC5439 Length: 282 February 11, 2000 15:51 Type: P Check: 934 ..

1 MDTVRIAVVG AGVGLSTAV CISKVPGCS ITVIDKFTF ETTSDVAAGM
 51 LIPIPTDPTP IHTOKWFRE TFNHLFAIAN SAEADAGVH LVSGIKSGG
 101 WTLTRIEDL WEHPSFDIV VNCGLGSRQ LAGDSKIFPV RGQVLOVAP
 151 WVEHFIRDS GLTIYIPGTS HTVLGTRQK GDMNLSPEAE NSREILSRCC
 201 ALPSLHGAC NIREKVLNP YRPGVRLQTE LARDGRLPV VHHYHGSGG
 251 GISVHWGTAL EEARLVSECV HALRTPIPKS NL
 !!AA_SEQUENCE 1.0
 P1:S25422 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 1 - Japanese quail mitochondrion (SGC1) (Fragment)
 C:Species: mitochondrion Coturnix coturnix japonica (Japanese quail)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Aug-1999
 C:Accession: S25422
 R:Desjardins, P.; Morais, R.
 J. Mol. Evol. 32, 153-161, 1991
 A:Title: Nucleotide sequence and evolution of coding and noncoding regions of a quail mitochondrion genome.
 A:Reference number: S25422; MUID:91178819
 A:Accession: S25422
 A:Molecule type: DNA
 A:Residues: 1-72 <DBS>
 A:Cross-references: EMBL:X57246; NID:912849; PIDN:CAA0522.1; PID:q12850
 C:Genetics: NDI
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 1
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; oxidoreductase; respiratory chain

S25422 Length: 72 February 11, 2000 15:51 Type: P Check: 6471 ..

1 SFLSPSELF SITLAKVLL IJSSFLWIRA SYPRFRYDQL MHLMLKNPLP
 51 LTIAMCLMHT SMPISYAGLP PA
 !!AA_SEQUENCE 1.0
 P1:S08622 - hydrogenase (EC 1.18.99.1) 3 chain 4 - Escherichia coli

N: Alternate names: formate hydrogenlyase chain 4; hydrogenase-3 protein D
 C: Species: *Escherichia coli*
 C: Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 26-Aug-1999
 C: Accession: S08622; F65052
 R: Boehm, R.; Sauter, M.; Boeck, A.
 Mol. Microbiol. 4, 231-243, 1990
 A: Title: Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenlyase components.
 A: Reference number: S08619; MUID:90251163
 A: Accession: S08622
 A: Molecule type: DNA
 A: Residues: 1-307 <BOE>
 A: Cross-references: EMBL:X17506; NID:9556890; PID:941683
 R: Blatter, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A: Title: The complete genome sequence of *Escherichia coli* K-12.
 A: Reference number: A64720; MUID:97426617
 A: Accession: F65052
 A: Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-307 <BLAT>
 A: Cross-references: GB:AE000356; GB:U00096; NID:92367153; PID:91789077; UMG:B2722
 A: Experimental source: strain K-12, substrain MG1655
 C: Genetics:
 A: Gene: hycD
 A: Map position: 58-59 min
 C: Superfamily: NADH dehydrogenase (ubiquinone) chain 1
 C: Keywords: oxidoreductase; transmembrane protein
 S08622 Length: 307 February 11, 2000 15:51 Type: P Check: 675 ..

1 MSVLPILQIA LVFAVAPLL SGITRVARAR LHNRRGPGVL QETDRIKLL
 51 GROSQGPDS GWFRLTPY MYGVALTIAT ALPVTGSP LPQLDLYTL
 101 LYLFAIAFF FAISGLDTGS PETAIGASRE AMGLVLEPM LILGLWAAQ
 151 VAGSTNISNI TDTVHMPIS GSPLVLC ACAPATRIEM GLPFDLAEV
 201 EQELQEGPLS EYSGSGGVM KMGISLQKLV VLOMFVGVFI PMGQETFTA
 251 GGLLALVIA IVKLVGVLV IALFENSMAR LRDDITPRIT MAGFPAFLA
 301 FVSLIAA

1:AA_SEQUENCE 1.0
 P1:T11338 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - aardvark
 C: Species: *mitochondrion Oryzteropus afer* (aardvark)
 C: Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 13-Aug-1999
 C: Accession: T11338
 R: Arnason, U.; Gullberg, A.; Janke, A.
 Proc. R. Soc. Lond. B Biol. Sci. 266, 339-345, 1999
 A: Title: The mitochondrial DNA molecule of the aardvark, *Oryzteropus afer*, and the position of the tubulin gene in the eutherian tree.
 A: Reference number: Z17633; MUID:99197468
 A: Accession: T11338
 A: Status: Preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-347 <ARN>
 A: Cross-references: EMBL:X18475; NID:e1427536; PID:e1427538; PIDN:CAB41622.1
 C: Genetics:
 A: Genome: mitochondrion
 A: Genetic code: SGC1
 A: Note: NADH2
 C: Superfamily: NADH dehydrogenase (ubiquinone) chain 2
 C: Keywords: mitochondrion; NAD; oxidoreductase
 T11338 Length: 347 February 11, 2000 15:51 Type: P Check: 2220 ..

1 NMPLFLMIM MTLSGTIT MLSHPTAW MGLMNNFAL IPIIKNHP
 51 RSEIATYF LTOATASML MIALNLNL SGMSTISIP SFTSFLIVT
 101 ALIMKLGAP EHEWLEPVTQ GTHLSSGLLI LTMOKLAPS ILYQISNIN
 151 PTLMTSAFL SILIGNGI NOTQRLKLA YSSIAHNGW ANILYINPTL
 201 TLNLLIYIT LLSAFMLI INYSTSTSL SLIMNKTPLI SLALNAILLS
 251 MGLPPLSGF IPKWLMOEL TKNLNAILP LMATLALNL YFYMRLIYSS
 301 SLTIIPSMNN MKIKMFNNH NIYGISDPM LLSMTILPLT PMYMSLF

1:AA_SEQUENCE 1.0
 P1:T11159 - NADH dehydrogenase 3 - hardbacked tick (*Rhipicephalus sanguineus*) mitochondrion
 C: Species: *mitochondrion Rhipicephalus sanguineus*
 C: Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
 C: Accession: T11159
 R: Black IV, W.C.; Roehrdanz, R.L.
 Mol. Biol. Evol. 15, 1772-1785, 1998
 A: Title: Mitochondrial gene order is not conserved in arthropods: prostratate and metastriate tick mitochondrial genomes.
 A: Reference number: Z17252; MUID:99083443
 A: Accession: T11159
 A: Status: Preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: DNA
 A: Residues: 1-114 <BLA>
 A: Cross-references: EMBL:AF081829; NID:94164556; PID:94164563; PIDN:AAD05523.1
 C: Genetics:
 A: Genome: mitochondrion
 C: Superfamily: NADH dehydrogenase (ubiquinone) chain 3
 C: Keywords: mitochondrion
 T11159 Length: 114 February 11, 2000 15:51 Type: P Check: 979 ..

1 MIFYLHITIF LVCLIMLNF FSLGFGKKA KKNNSPECG FDFSLSRVP
 51 FSLKEFEVGI VFLLFVEIV VILPPLVMM TKNLMEVFSF TFINPLIVAG
 101 LLYEFKYSML DRLK

1:AA_SEQUENCE 1.0
 F1:S42244 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Anser caerulescens mitochondrion (SGC1) (fragment)
 C: Species: *mitochondrion Anser caerulescens*
 C: Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
 C: Accession: S42244
 R: Quinn, T.W.; Wilson, A.C.
 J. Mol. Evol. 37, 417-425, 1993
 A: Title: Sequence evolution in and around the mitochondrial control region in birds.
 A: Reference number: S42243; MUID:94141942
 A: Accession: S42244
 A: Molecule type: DNA
 A: Residues: 1-214 <QU>
 A: Cross-references: EMBL:X75772; NID:9450491; PIDN:CAA53393.1; PID:91164899
 C: Genetics:
 A: Gene: ND5
 A: Genome: mitochondrion
 A: Genetic code: SGC1
 C: Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C: Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; oxidoreductase; respiratory chain
 S42244 Length: 214 February 11, 2000 15:51 Type: P Check: 1655 ..

1 GSVLDPSLN TSYLNTALL LTLMAFAFA TYSIMTILV OAGOTRIPPM
 51 VSMNNENPLI TAPILRLAG SITAGMIIS FITPKTPPM TMDLITRTAA
 101 ILMTITGITL ALELSNMTH LTYKPNPLM NSSALGYFN PLVHREFSKN

151 LTERGNIAL HLIDSLWKK MGPEGLAELQ VAASKATSM HTGLIKAYLS
201 SFALSTIVMI LMTM

11AA_SEQUENCE 1.0
PI: F64999 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain J -
Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 20-Aug-1999
C:Accession: F64999; S38310; S37067
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; M01D:97426617
A:Accession: F64999
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-184 <BLAT>
A:Cross-references: GB:AE000317; GB:U00096; NID:91788605; PIDN:AMC75340.1;
PID:91788616; WMC:P.b2280
A:Experimental source: strain K-12, substrain MG1655
R:Weidner, U.; Geier, S.; Plock, A.; Friedrich, T.; Lell, H.; Welts, H.
J. Mol. Biol. 233, 109-122, 1993
A:Title: The gene locus of the proton-translocating NADH:ubiquinone
oxidoreductase in Escherichia coli. Organization of the 14 genes and
relationship between the derived proteins and subunits of mitochondrial complex
I
A:Reference number: S38310; M01D:93389724
A:Accession: S38310
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-84; T, 86-125, 'R', 127-184 <ME2>
A:Cross-references: EMBL:X68301; NID:9444012; PIDN:CAA6369.1; PID:9397907
C:Experimental source: strain AN387
C:Genetics:
A:Gene: nuoJ
A:Map position: 49.5 min
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6
C:Keywords: membrane-associated complex; NAD; oxidoreductase; transmembrane
protein
F64999 Length: 184 February 11, 2000 15:51 Type: P Check: 2909 ..

1 MEFAFYICGL IALILATLRI THTNPVHAL YLITSLAIS GVEFSLGAYF
51 AGALEITIVA GAIMVLFVR VMNLNGSE IEGERQWLR QVMIGPAIIS
101 AIMLVIVYA ILGVNDQGD GTPISAKAVG ITLFGPVYLA VELASMLLIA
151 GLVYAFHVR EERAGEVLSN RKDSAKRKT EEHA

11AA_SEQUENCE 1.0
PI: S52970 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 6 - honeybee
mitochondrion (SGC4)
C:Species: mitochondrion Apis mellifera (honeybee)
C:Date: 15-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C:Accession: S52970
R:Crozler, R.H.; Crozier, Y.C.
Genetics 133, 97-117, 1993
A:Title: The mitochondrial genome of the honeybee Apis mellifera: complete
sequence and genome organization.
A:Reference number: S52960; M01D:93114603
A:Accession: S52970
A:Molecule type: DNA
A:Residues: 1-167 <CRO>
A:Cross-references: EMBL:L06178; NID:9336279; PIDN:ABB96808.1; PID:9829009
C:Experimental source: Ligustica
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC4

1 MMLTIMSK IMFSLISMI LTIYNIFN SPSULLIYLI SYSIYSLWM
51 FTGSMNSL IMMLIVELS GMLMFYFI SLINPEPLK MPEIOTLRL
101 IITMKIYK LSONEHYFN FNIDIMLYL MKMNSTLEFI MLMLITLIL
151 LMTKITYEK KTLRKK

11AA_SEQUENCE 1.0
PI: S78138 - cytochrome-c oxidase (EC 1.9.3.1) chain III - Reclinomonas
americana (ATCC 50394) mitochondrion
C:Species: mitochondrion Reclinomonas americana
A:Variety: ATCC 50394
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Aug-1999
C:Accession: S78138
R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux,
C.; Sankoff, D.; Turmel, M.; Gray, M.W.
Nature 387, 493-497, 1997
A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in
miniature.
A:Reference number: S78127; M01D:97311393
A:Accession: S78138
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-267 <LAN>
A:Cross-references: EMBL:AF007261; NID:92258325; PIDN:AD11871.1; PID:92258337
A:Experimental source: ATCC 50394
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
1997
C:Genetics:
A:Gene: cox3
A:Genome: mitochondrion
C:Superfamily: cytochrome-c oxidase chain III
C:Keywords: electron transfer; membrane-associated complex; mitochondrial inner
membrane; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory
chain; transmembrane protein
F:4-18/Domain: mitochondrial matrix #status predicted <MM1>
F:19-37/Domain: transmembrane #status predicted <TM01>
F:38-45/Domain: intracristal #status predicted <TM02>
F:46-71/Domain: transmembrane #status predicted <TM03>
F:72-77/Domain: mitochondrial matrix #status predicted <MM2>
F:78-110/Domain: transmembrane #status predicted <TM04>
F:111-133/Domain: intracristal #status predicted <ITC2>
F:134-157/Domain: transmembrane #status predicted <TM05>
F:158-160/Domain: mitochondrial matrix #status predicted <MM3>
F:161-188/Domain: transmembrane #status predicted <TM06>
F:189-195/Domain: intracristal #status predicted <ITC3>
F:196-228/Domain: transmembrane #status predicted <TM07>
F:229-237/Domain: mitochondrial matrix #status predicted <MM4>
F:238-261/Domain: transmembrane #status predicted <TM08>
F:262-266/Domain: intracristal #status predicted <ITC4>
S78138 Length: 267 February 11, 2000 15:51 Type: P Check: 8115 ..

1 MSQTFVKRHP YHVDOSPWP LITSIGTICS TFGVWFHS YPNGFIALL
51 GVSTILFSLY AMCRDIYREG TYOGHTAAV QNGLRIGMIL FITSEYMFV
101 SFVAFHSS LSPITIEIGAV WPQCIETLN AMDVPFNTV ILMGATVAT
151 WSHHMHGNG RTQSIIGLIF TIIAVTFEG LQVNEYREAS FSIADGIYGS
201 TFVYATGFHG FHVIVGCMU SYCLVRELY HFTTHHGF EASAWMHFV
251 DVVWLFLETT IYWGNG

```

!!AA_SEQUENCE 1.0
P1:S07557 - cytochrome-c oxidase (EC 1.9.3.1) chain III - bracket fungus
C:Species: Schizophyllum commune mitochondrion
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Aug-1999
C:Accession: S07557
R:Phelps, L.G.; Burke, J.M.; Ullrich, R.C.; Novotny, C.P.
Curr. Genet. 14: 401-403, 1988
A:Title: Nucleotide base sequence of the mitochondrial COIII gene of
Schizophyllum commune
A:Reference number: S07557; MUID:89106272
A:Accession: S07557
A:Molecule type: DNA
A:Residues: 1268 <PHE>
A:Cross-references: EMBL:X15748
C:Genetics:
A:Gene: COIII
A:Genome: mitochondrion
C:Superfamily: cytochrome-c oxidase chain III
C:Keywords: electron transfer; membrane-associated complex; mitochondrial inner
membrane; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory
chain; transmembrane protein

S07557 Length: 268 February 11, 2000 15:51 Type: P Check: 7659 ..

1 MTSIKYQSF SAHVOHSPW PILVSFLFN LAIGVLTWH GYSHSTTFD
51 LGLAVTQSI LMTDRIVIE GSFGLDHTKQ VQGLIIGFI LFIISVFAF
101 ISVMAFVHS ALSPAVELGS TWPVGIIP LDFSLPLFWI ILLSGAFV
151 TYGHHAFSG KRDSIIGLF LVALALIFS YFOAFEXIHA GFSMSDVFQ
201 TVFASVGLH GIHMIGTLF LFVSELRQVN YQTKRHNGJ LETSILWHE
251 VDLWLEFL VYFNGCA

!!AA_SEQUENCE 1.0
P1:T09997 - lipoygenase (EC 1.13.11.12) - southern Asian dodder (fragment)
C:Species: Cuscuta reflexa (southern Asian dodder)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T09997
R:Subramaniam, K.; Mahadevan, S.
Submitted to the EMBL Data Library, January 1994
A:Description: Cloning and characterization of a cDNA coding for lipoygenase
associated with cytokinin-induced haustoria formation in Cuscuta reflexa.
A:Reference number: Z16913
A:Accession: T09997
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1385 <SUB>
A:Cross-references: EMBL:U05041; NID:9450230; PID:9450231
C:Genetics:
A:Gene: lox
A:Function:
A:Description: hydroperoxidation of unsaturated fatty acids
A:Superfamily: lipoygenase
C:Keywords: fatty acid oxidation; oxidoreductase

T09997 Length: 385 February 11, 2000 15:51 Type: P Check: 1856 ..

1 WMESHSKSC FAIPQNVIA WMKTRAKAHV LAHDSGVHOL VSHMLRTCA
51 IEPYIIASNR QLSMLPTIFR LKPRFRITM EINNLAERAL INANVITST
101 FSPGKSHML SSMAVDEMG FDLQALPADL ISRGLAEDP TAPHGLKLA
151 EYPRANDGL VMDIIEWY TDYVTHYPE SSRVESTEL QEMMKELITV
201 GHADRDAGG WPDKTPEDL IGLITTIWV CSQHNAVNF GQYVDAQFP
251 NRPITARTM PTHDSESEW EHPMKPEDT LKCFPSQLQ ARIKVAVIDV

301 LSNHSPDEEY LCASPEYWI DEPTINAEF RESGCFELE GIIDGRNDR
351 NLNRRNGAGV LPYELIKRPS GPGVYKGV NPISLI

!!AA_SEQUENCE 1.0
P1:S43771 - phosphatidylcholine desaturase (EC 1.3.1.35) - Synecococcus sp.
(strain PCC 7002)
N:Alternate names: Delta-12 desaturase; fatty acid desaturase (EC 1.14.99.-)
(Misidentification)
C:Species: Synecococcus sp.
A:Variety: PCC 7002
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 28-May-1999
C:Accession: S43771
R:Sakamoto, T.; Wada, H.; Nishida, I.; Ohmori, M.; Murata, N.
Plant Mol. Biol. 24: 643-650, 1994
A:Title: Identification of conserved domains in the Delta-12 desaturases of
cyanobacteria.
A:Reference number: S43770; MUID:94207189
A:Accession: S43771
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1347 <SAX>
A:Cross-references: GB:D13779; NID:9488510; PIDN:BA002922.1; PID:dl003428;
PID:9488511
C:Superfamily: omega-3 fatty acid desaturase
C:Keywords: oxidoreductase

S43771 Length: 347 February 11, 2000 15:51 Type: P Check: 2848 ..

1 MTSVTRPSA TLLEKHPNL RLRLDILTLP RSYVEINPLK AMSVLLSVA
51 AVGCYALIA IAPWYLLPV WFLTGTLTG FVYIGHDGH RSFRKKNVN
101 NLVGHIAFLP LIYFHSWRI LNNHHRXTN NMEDNANAP FTPELYDSDP
151 AFIRAVYRAI RGLIMWIASV IHOULKHEW FAFEGKORBO VRFSALEVVI
201 AGAIFPVNF YTLGVWGVYK FWLMPWLGTH FWSFTFLVH HTVEIDPFSY
251 RKNWEALAQ LSGTVACDYP KWVEVLCHDI NVHVPVHLST GIPSYNLAKA
301 VASIKQNGE YLVETKFSWE LMKATTEGCH LYDAENHYIS FAOHOKR

!!AA_SEQUENCE 1.0
P1:S74692 - trna(m137)methyltransferase - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein s111198
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 18-Jun-1999
C:Accession: S74692
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
Miyajima, N.; Hirosewa, M.; Sugiyura, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpou, S.;
Takenuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3: 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium
Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and
assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S74692
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1231 <RNA>
A:Cross-references: EMBL:D90901; GB:AB001339; NID:91651897; PIDN:BA16843.1;
PID:dl017576; PID:91651917
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
1996
C:Genetics:
A:Gene: trmD
A:Start codon: GTG
C:Superfamily: tRNA (guanine-N1) methyltransferase
C:Keywords: methyltransferase

```

```

S74692 Length: 231 February 11, 2000 15:51 Type: P Check: 6210 ..

1 MOPDVLTLP DFETSPLOSG LLGKALEKAI ASYNLINPRD FTTDKHRRVD
51 DEPYGGGVGM VIKPEPIFAA VESLPVLSKR EVILMPOGO PMDQALFREL
101 TNYDQVLVLC GHYEGVDREV COLVTRVSL GDFVLTCGEI PALTLINGVI
151 RLPLPGTVGKE ASLIAESFST DLDLDPHYTR PPFVRLGLAVP PVLISGNHOA
201 IAQWRLEQGE ERTQQRRLPD WQKWQDRQPS P

!!AA_SEQUENCE 1.0
PI:A71119 - Probable aspartate carbamoyltransferase catalytic chain -
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 18-Jun-1999
C:Accession: A71119
R:Kawabuchi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohkuni, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: A71119
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <RAW>
A:Cross-references: GB:AP000003; NID:93236130; PIDN:BA29811.1; PID:dl030754;
PID:g3257128
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by
GenBank
C:GeneID: P07070
A:Gene: P07070
C:Superfamily: ornithine carbamoyltransferase: aspartate/ornithine
carbamoyltransferase homology
F:6-305/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
A71119 Length: 308 February 11, 2000 15:51 Type: P Check: 9980 ..

1 MEMKGRDVIS IDFSKEDIE VILSTARLE KEMKEKGQLE YAKGKIATL
51 FEFPSTRTRL SPESAMHRLG GSVIGFAEAS TSSVKKGESL RDTIKIVEQY
101 SDVIIVIRHPK EGARLALEV ADIPVINAGD GSNQHTPTOL LDLYTIKKEF
151 GTIDGLKIGL LGDLKYGRTV HSLAEALAFY DVELYIISPE LLRMPKHIVE
201 ELNRGKRTV ETTKLEEVIG ELVDLYVTRI OKERFDEDE YLAKXGSGVY
251 NKLILENVKD SLRIMHPLPR VDEIHPEVDK TKHAIYFKOV FNGVPVRNAL
301 LALVLCVI

!!AA_SEQUENCE 1.0
PI:A70959 - Probable aspartate carbamoyltransferase - Mycobacterium
tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-Aug-1999
C:Accession: A70959
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: C70815
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-403 <COI>
A:Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PID:el254001;
PID:g2916918
A:Experimental source: strain H37Rv
A:Gene: fadA
C:Superfamily: acetyl-CoA acetyltransferase
C70815 Length: 403 February 11, 2000 15:51 Type: P Check: 556 ..

1 MSEAFIYDA IRTPRGOKN GSLHEVKKPS LVVGLIDELR KRRPDIDENTL
51 ISDVIIGCVS PVGOGGIDA RAAYLVASGMP VTSGGVQLNR FCASGLEAVN
101 TAAQXKRSWM DDVLVAGSVE SMSRVPMSGD GGAGGLDPAT NYVPMFVPOS
151 IGDLDLATIE GFSREDPDAY ALRSQKAAE AMSGTYFAKS VVYVRQNGL
201 LIIDHEHNR PDTTKRGLK LKPAFGLAA LGFDDVALQ KYHWKXINH

```

```

251 VHTGNSSGI VDGALVIG SAAGKLOGL TPRARIATA TSGADPIML
301 TGPTATKRV LDRAGLTVD IDLFELNEAF ASVYLFQKD LNIPEKLVY
351 NGGALAMGP LGATGAMITG TWDLELRN ARRALITICI GGMGVATII
401 ERV

!!AA_SEQUENCE 1.0
P:112981 glucuronosyl transferase-like protein - Arabidopsis thaliana
N:Alternate names: protein T6H20.280
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 20-Sep-1999
C:Accession: 112981
R:Choliste, N.; Robert, C.; Brottier, P.; Wincker, P.; Gattolico, L.;
Artiguenave, F.; Saulin, W.; Weissenbach, J.; Mevius, H.W.; Mayer, K.F.X.;
Lemcke, K.; Schueller, C.; Queller, F.; Salanoubat, M.
Submitted to the Protein Sequence Database, July 1999
A:Reference number: 217586
A:Accession: 112981
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <CHO>
A:Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.280
A:Experimental source: cultivar Columbia; BAC clone T6H20
C:Genetics:
A:Gene: ATSP:T6H20.280
A:Map position: 3
A:Introns: 161/1
C:Superfamily: flavonol O3-glucosyltransferase

T12981 Length: 452 February 11, 2000 15:51 Type: P Check: 8558

1 MEKRVKRII VLVPAAGH VTPMQLGKA LQSKGLITV AQRFQIGS
51 SIQHFGDF VTIPESLPQS ESKKLPAEV LMNNTKSEA SFECISQIS
101 MOGNDIACI IYDKIMFCE AAKKERIPS VIFSSSATI QVCYCLSEL
151 SAEFLIDMK DPEKDKYLE GLHPLRYKDL PTSGPPLPEP LLMKREYVN
201 KRASAVITN TASCLESISL SWIQOELGIP VYPLGLHIT ASSPGSILQ
251 EDMSCIEWIN KOKPRSVIYI SLGTAKMET KEMLEMAGL LNSNOPIMV
301 IRPGSVAGEF WIELPEEVI KMYTERGYIA KPAQIEVLG HPVAGGFWSH
351 CGNNTLESII VEGVPMICRP LQGEOKINAM VIESYWKIGI QLEGEVEREG
401 VERAVKRLII DEGAAMRER ALDLKERLNA SVRSGSSSYN ALDELVKFLN
451 TE

!!AA_SEQUENCE 1.0
P:1140986 streptomycin/spectinomycin adenylyltransferase - Escherichia coli
C:Species: Escherichia coli
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 18-Jun-1999
C:Accession: 140986
R:Hollingshead, S.; Vapnek, D.
Plasmid 13, 17-30, 1985
A:Title: Nucleotide sequence analysis of a gene encoding a
streptomycin/spectinomycin adenylyltransferase.
A:Reference number: 140986; MUID:85191077
C:Accession: 140986
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-323 <RES>
A:Cross-references: EMBL:X02340; NID:940880; PIDN:CAA26199.1; PID:9581035
C:Superfamily: streptomycin 3'-adenylyltransferase

140986 Length: 323 February 11, 2000 15:51 Type: P Check: 7621

1 MTAQNRFSM LVMTCFFGV OSNPRASKOO ARVAVGCLM LMSNDVTOQ
51 GSRPKTKINI MREAVIAEVS TOLSEVGYI ERHLEPTLLA VHLGSAVDG
101 GLKPHSIDL LVTVVRIDE TTRRALINDL LETSASGES EILRAVEYTI
151 VVHDDIIPWR YPAKRELQFG EMORNDILAG IFEPATIDID LALLTKARE
201 HSAVLGPAA EELFDPVEQ DFEALNETL TLWNSPPDMA GDERNVLTLL
251 SRIWYAVTG KIAPDVAD WAMERLPAQY QPVILEAROA YLGOEDRILA
301 SRADQLEERY HYKGEITKV VGR

!!AA_SEQUENCE 1.0
P:1165760 alcohol sulfotransferase (EC 2.8.2.2) - rat
N:Alternate names: hydroxysteroid sulfotransferase; senescence marker protein 2B
C:Species: Rattus norvegicus (Norway rat)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 18-Jun-1999
C:Accession: 165760; 152407
R:Matabe, T.; Ogura, K.; Satsukawa, M.; Okuda, H.; Hiratsuka, A.
Chem. Biol. Interact. 92, 87-105, 1994.
A:Title: Molecular cloning and functions of rat liver hydroxysteroid
sulfotransferases catalyzing covalent binding of carcinogenic polycyclic
arylmethanols to DNA.
A:Reference number: 152849; MUID:94306585
A:Accession: 165760
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-285 <RES>
A:Cross-references: GB:D14989; NID:9303801; PIDN:BAA03634.1; PID:9303802
R:Song, C.S.; Kim, J.M.; Roy, A.K.; Chatterjee, B.
Biochemistry 29, 542-551, 1990
A:Title: Structure and regulation of the senescence marker 2 gene promoter.
A:Reference number: 152408; MUID:90148982
A:Accession: 152407
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 2-68 <RES2>
A:Cross-references: GB:M29302; NID:9206995; PIDN:AAA42152.1; PID:9206996
C:Genetics:
A:Introns: 46/1
C:Superfamily: alcohol sulfotransferase
C:Keywords: sulfotransferase

165760 Length: 285 February 11, 2000 15:51 Type: P Check: 800

1 MMSDYTFEG IPPPAFWFSK EILENSCKR VYKEDDLITL TYKSGTNWL
51 IEIVCLIQTK GDPKWIQSNP IWDRSPWIEF GSGYDKLTKM EGPRLMTSHL
101 PMHLESKSLF SSKAKVIYLI RNPRDVLVA YFFWSKIALE KKPDSIGTIY
151 EMFLKGNVAY GSWFEHINQW LSMREWDNPL VLYYEDMKMD TWGSIKICD
201 FLGKLEPDE LNLVLYKSSF QVYKNNMSN YSLMEKEIIL TGFTFMKGT
251 TNDWKNHFTV AQAEAFDKVF QEKWAGPPPG MEPPWE

!!AA_SEQUENCE 1.0
P:1152265 dual specificity phosphatase (EC 3.1.3.-) 1 - rat
N:Alternate names: protein-tyrosine-phosphatase CL100;
protein-tyrosine-phosphatase, nonreceptor type 10
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 22-Jun-1999
C:Accession: 552265
R:Muda, M.; Schlegel, W.; Arkininstall, S.
Submitted to the EMBL Data Library, January 1995
A:Description: Pathways regulating CL100 gene expression in pituitary cells.
A:Reference number: 552265
A:Accession: 552265
A:Status: preliminary

```

A:Molecule type: mRNA
 A:Residues: 1-367 <MCD>
 A:Cross-references: EMBL:X84004; NID:g642264; PIDN:CA58828.1; PID:g642265
 C:Superfamily: dual specificity phosphoprotein phosphatase 1; VHL-type dual
 C:Specificity phosphoprotein phosphatase homology
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase
 F:181-312/Domain: VHL-type dual specificity phosphoprotein phosphatase homology
 <VHL>
 F:258/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:266/Binding site: substrate phosphate (Arg) #status predicted
 S52265 Length: 367 February 11, 2000 15:51 Type: P Check: 6443 ..

1 MMEVGIIDA GGLRALLER AQCLLDOR SFFAFMGHI VGSVNRFT
 51 IVRRRAKAM GLEHIVPTE LRGRLAGY HAVVLDEBS ALDCAKRG
 101 TLALAGALC REARSTQVFF LGGYEAFSA SCPELCKOS TPNGLSLPIS
 151 TSVPDSASG CSSCSTPLYD OGCPYEILSF LYIGSAYHAS RDMDLALCI
 201 TALINVSANC PNHEGHYQY KSIPYEDNHK ADISSMNEA IDFIDSINDA
 251 GGRFVHCOA GISRSATIL AYLMRTNRVK LDEAFEFVKO RRSIISPNS
 301 FMGQLAPES QVLAPHCAS ACSPMAVLD RGTSTTVFN FVSIPIVHPT
 351 NSALNYIQSP ITTSPSC

!!AA_SEQUENCE 1.0
 P1:TI2078 - ribonuclease (EC 3.1.-.-) - almond
 C:Species: Prunus dulcis (almond)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
 C:Accession: TI2078
 R:Ushijima, K.; Sassa, H.; Tao, R.; Yamane, H.; Dandekar, A.; Gradziel, T.; Hirano, H.
 M.Ol. Gen. Genet. 260, 261-268, 1998
 A:Title: Cloning and characterization of cDNAs of the S-RNases in almond (Prunus dulcis): primary structural features and sequence divergence of the rosaceae S-RNases.
 A:Reference number: 217403
 A:Accession: TI2078
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-231 <USH>
 A:Cross-references: EMBL:AB011469; NID:d1241928; PID:d1035644
 A:Experimental source: strain Mission, pistil
 C:Superfamily: Enterobacter ribonuclease
 C:Keywords: hydrolase
 TI2078 Length: 231 February 11, 2000 15:51 Type: P Check: 3061 ..

1 MAMKSSAF LVAFAFLC FIMSTGSIVY FQVQOMPPT NCVARKKRC
 51 SNRPLOQFT IHGIMPSNTS NPTKSCNCG SQNFPTVSP KMRVYKRRM
 101 PDVESGNDTR FWEQWNRKG TCSESSLNOM QYFERSHEW YSFNTEILK
 151 NASIVPHPTQ TWKYSIDVAP IKTAKRPV LRCRPDAON KSGPKTOLH
 201 EVFCEYEA LKQIDCNRTA GCMNVDIK Q

!!AA_SEQUENCE 1.0
 P1:B71518 - probable lipoprotein signal peptidase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 26-Aug-1999
 C:Accession: B71518
 R:Seebens, R.S.; Kaiman, S.; Lammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.S.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.

A:Reference number: A71570; MUID:99000809
 A:Accession: B71518
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-167 <ARN>
 A:Cross-references: GB:AE001314; GB:AE001273; NID:g3328833; PID:g3328836
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: lspa
 C:Superfamily: lipoprotein signal peptidase
 B71518 Length: 167 February 11, 2000 15:51 Type: P Check: 3917 ..

1 MPTSLPTFL TLULLASIDW VSKLVVLKS CQSPSSSAF LXYWGHFS
 51 FLIPSENEG AAFGLFTQYK IPLIFRCV ILGLALFLRI KYSLHRTIR
 101 VALTIIAGA LGNVGDILY GRVDFLSLS YSWRPFSFN LADAFISIGT
 151 LLLIGHLYFT KSKKYE

!!AA_SEQUENCE 1.0
 P1:S26429 - probable dUTP pyrophosphatase (EC 3.6.1.23) - fowl adenovirus 1
 C:Species: Aviadenovirus gall (fowl adenovirus 1, CELO)
 C:Date: 07-May-1993 #sequence_revision 02-Aug-1994 #text_change 12-Apr-1995
 C:Accession: S26429
 R:Akopian, T.A.; Kaverina, E.N.; Naroditsky, B.S.; Tikhonenko, T.I.
 Submitted to the EMBL Data Library, October 1992
 A:Description: Avian adenovirus CELO: sequence of the DNA fragment (92-100%).
 A:Reference number: S26428
 A:Accession: S26429
 A:Molecule type: DNA
 A:Residues: 1-178 <AKO>
 A:Cross-references: EMBL:Z17216
 C:Superfamily: retroviral proteinase
 C:Keywords: hydrolase
 S26429 Length: 178 February 11, 2000 15:51 Type: P Check: 9885 ..

1 MDPFGSSVPC PSTSDLPKP KLYVRUSPH APPVPRATNG AAGYDLFSAY
 51 DIKVPARGRA LVPTDLVQF PGCGYRIAP RSGIAKFTI DVGACVIDPD
 101 YRGNVSVLF NSESSFNIR RGRVAVQLIL ERIMPELSE LTOLGERTDRG
 151 ASGFGFTGKG AVDRNQRVYL EWLTPGSR

!!AA_SEQUENCE 1.0
 P1:S42592 - urease (EC 3.5.1.5), tissue-ubiquitous - soybean (fragment)
 C:Species: Glycine max (soybean)
 C:Date: 19-Mar-1997 #sequence_revision 24-Jul-1997 #text_change 07-May-1999
 C:Accession: S42592
 R:Totterly, R.S.; Griffin, J.D.; Yenofsky, R.L.; Polacco, J.C.
 M.Ol. Gen. Genet. 242, 404-414, 1994
 A:Title: A single gene (Eud) encodes the tissue-ubiquitous urease of soybean.
 A:Reference number: S42592; MUID:94166749
 A:Accession: S42592
 A:Molecule type: mRNA
 A:Residues: 1-131 <TOR>
 A:Cross-references: EMBL:S69179
 C:Genetics:
 A:Gene: Eud
 C:Superfamily: urease; urease 11k chain homology; urease 12k chain homology; urease 62k chain homology
 C:Keywords: hydrolase
 F:1-114/Domain: urease 62k chain homology (fragment) <U62>
 S42592 Length: 131 February 11, 2000 15:51 Type: P Check: 4170 ..

1 PSFEGAPEM VIKGVVANA DMGDPNASIP TPPEVAKRPM FGLCKAKGA
 51 LSIAPVSKAA VDORVHALYG LNKRYKAVGN VRKRLTLDK LNDSPQITV


```

101 DPNKVTAD GEVLTSFAT FVPLSRNYFL F
11AA_SEQUENCE 1.0
P1:151663 - arginase 1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151663
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151663; MUID:95014323
A:Accession: 151663
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008406; NID:9497227; PIDN:AAA56891.1; PID:9497228
C:Superfamily: arginase 1

15163 Length: 360 February 11, 2000 15:51 Type: P Check: 4321 ..

1 MSIRSNFVRL LKKOVSIIKL OKKCSHSAV IGAPESKQK RRGVEHGPAA
51 IRSAGLIERL SNLGNVCFD GDHFSQVFN DELYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCYTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NPLTPSGN LHGQPSFLL RELQDKVPI PGFSMAKPC LSKSDIYIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEHNHGM LSAVDLVEVN
301 PVLAITSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:151664 - arginase 2 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151664
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151664; MUID:95014323
A:Accession: 151664
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008407; NID:9497229; PIDN:AAA56892.1; PID:9497230
C:Superfamily: arginase 1

15164 Length: 360 February 11, 2000 15:51 Type: P Check: 4641 ..

1 MSIRSNFVRL LKKOVSIIKL OKKCSHSAV IGAPESKQK RRGVEHGPAA
51 IRSAGLIERL SNLGNVCFD GDHFSQVFN DELYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCYTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NPLTPSGN LHGQPSFLL RELQDKVPI PGFSMAKPC LSKSDIYIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEHNHGM LSAVDLVEVN
301 PVLAITSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:151665 - arginase 3 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Jun-1999
C:Accession: 151665
R:Paterson, D.; Shi, Y.B.
J. Biol. Chem. 269, 25328-25334, 1994
A:Title: Thyroid hormone-dependent differential regulation of multiple arginase
genes during amphibian metamorphosis
A:Reference number: 151665; MUID:95014323
A:Accession: 151665
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-360 <PART>
A:Cross-references: EMBL:008408; NID:9497231; PIDN:AAA56893.1; PID:9497232
C:Superfamily: arginase 1

15165 Length: 360 February 11, 2000 15:51 Type: P Check: 4724 ..

1 MSIRSNFVRL LKKOVSIIKL OKKCSHSAV IGAPESKQK RRGVEHGPAA
51 IRSAGLIERL SNLGNVCFD GDHFSQVFN DEOYSIVKH PRYGLACKY
101 LAEVSRAVG AGHTCYTLGG DSHLAFSGIT GHAQCCPDIC VIWDAHADI
151 NPLTPSGN LHGQPSFLL RELQDKVPI PGFSMAKPC LSKSDIYIGL
201 RDLDPAEQFI LKNYISYS MRHIDCMGIR KWEKTFDOL LGRDRPIHL
251 SFDDIAPDPA LAPATGPVI GGLTYREGV ITEEHNHGM LSAVDLVEVN
301 PVLAITSEV KATANLAVDV IASCFQTR EGAHTRADTII DVLPTPSTY
351 ESDNEQVRI

11AA_SEQUENCE 1.0
P1:J50609 - biastidicin-S deaminase (EC 3.5.4.23) - Bacillus cereus
C:Species: Bacillus cereus
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Aug-1999
C:Accession: J50609; PS0264
R:Kobayashi, K.; Kamakura, T.; Tanaka, T.; Yamaguchi, I.; Endo, T.
Agric. Biol. Chem. 55, 3155-3157, 1991
A:Title: Nucleotide sequence of the bar gene and N-terminal amino acid sequence
of biastidicin S deaminase from biastidicin S resistant Escherichia coli TK121.
A:Reference number: J50609; MUID:92144117
A:Accession: J50609
A:Molecule type: DNA
A:Residues: 1-140 <KOB>
A:Experimental source: strain K55-S1
A:Accession: PS0264
A:Molecule type: protein
A:Residues: 1-46 <KOB1>
C:Genetics:
A:Gene: bar
C:Superfamily: cytidine deaminase
C:Keywords: hydrolase

J50609 Length: 140 February 11, 2000 15:51 Type: P Check: 8425 ..

1 MTFNISOQD LELVEVATEK ITMLYEDNKH HVGAAITKT GELISAVHE
51 AYIGRTVCA EALIGSAVS NGQKDPDITV AVRHPSDEV DSRIRVSPC
101 GNCRELISDY APDCVLIEM NGKLVKTIE ELIPKXTRN

11AA_SEQUENCE 1.0
P1:S37304 - salm protein - Salmonella typhimurium (fragment)
C:Species: Salmonella typhimurium
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C:Accession: S37304
R:Goldsman, E.A.; Ochman, H.
EMBO J. 12, 3779-3787, 1993
A:Title: Cognate gene clusters govern invasion of host epithelial cells by
Salmonella typhimurium and Shigella flexneri.

```

A:Reference number: S37304; MUID:94008985
A:Accession: S37304
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-335 <GRO>
A:Cross-references: EMBL:X73525; NID:g404286; PIDN:CAAS1921.1; PID:g404287
C:Genetics:
A:Gene: spaL
C:Superfamily: H⁺-transporting ATP synthase alpha chain; H⁺-transporting ATP synthase alpha chain homology
C:Keywords: P-loop
F:63-70/Region: nucleotide-binding motif A (P-loop)
F:86-254/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>
S37304 Length: 335 February 11, 2000 15:51 Type: P Check: 2135 ..
1 DPGKIVIERF TPEVAPISE RVIDVAPPSY ASRVGAREPL ITGVRAIDGL
51 LTGCGGQRMG IFASAGCCGT MLMHMLEQT EADVFIYGLI GERGEVTEF
101 VDMIRASHKK EKCVLVFATS DPEVDRCNA AQLATTVAEX FRDQCKRYVL
151 FIDSMTRIVAR ALRDVALASG ERPARRGYPA SVFDNLPRLL ERPGATSEGS
201 ITAFYTVLLE SEEDADPMAD EIRSIDGHL YLSRKLAGOG HYPALDIVKS
251 VSRVFGQVTT PTHAEQASAV RKLMTREEL QLEFIDGEXR PGENDINDRA
301 MOKRSLKAW LCPVAQYSS FDDTLGKNA FADQN
11AA_SEQUENCE 1.0
P1:S01397 - H⁺-transporting ATP synthase (EC 3.6.1.34) chain I - thermophilic bacterium PS-3
C:Species: thermophilic bacterium PS-3
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 22-Jun-1999
C:Accession: S01397
R:Ohta, S.; Yohda, M.; Ishizuka, M.; Hirata, H.; Hamamoto, T.; Ohtawa-Hamamoto, Y.; Matsuda, K.; Kagawa, Y.
Biochim. Biophys. Acta 933: 141-155, 1988
A:Title: Sequence and over-expression of subunits of adenosine triphosphate synthase in thermophilic bacterium PS3
A:Reference number: S01397; MUID:88163679
A:Accession: S01397
A:Molecule type: DNA
A:Residues: 1-127 <OHT>
A:Cross-references: EMBL:X07804; NID:g45808; PIDN:CAA30647.1; PID:g45809
C:Superfamily: Bacillus H⁺-transporting ATP synthase chain I
C:Keywords: ATP biosynthesis; hydrolase
S01397 Length: 127 February 11, 2000 15:51 Type: P Check: 4392 ..
1 MGNLQAMFWR QVRILYLLA IYTGFGFTF YKTVFSLIL GTISILMKW
51 NLMTKIKFG QAVAAKKVR TLGLSLRAL AALAAIVLTF YQGFHIVPT
101 VLGLMSTYIV IIDFFHHK KNDLQA
11AA_SEQUENCE 1.0
P1:S22348 - H⁺-transporting ATP synthase (EC 3.6.1.34) delta chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S22348; T00493
R:Jordan, E.M.; Breen, G.A.M.
Biochim. Biophys. Acta 1130, 123-126, 1992
A:Title: Molecular cloning of an import precursor of the delta-subunit of the human mitochondrial ATP synthase complex.
A:Reference number: S22348; MUID:92182007
A:Accession: S22348
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-168 <OOR>
A:Cross-references: EMBL:X63422; NID:g12585; PIDN:CAA45016.1; PID:g12586
R:Jamerdin, J.E.; McCreedy, P.M.; Skowronek, E.; Adamson, A.M.; Burkhardt-Schultz, K.; Gordon, L.; Kyle, A.; Ramirez, M.; Stilaengen, S.; Phan, H.; Velasco, N.; Gaines, J.; Danganan, L.; Poundstone, P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Altix, C.; Andreise, T.; Trankelm, M.; Amico-Keller, G.; Coefield, J.; Duarte, S.; Lucas, S.; Bruce, R.; Thomas, P.; Qian, G.; Krommiller, B.; Arellano, A.; Montgomery, M.; Ow, D.; Nolan, M.; Trong, S.; Kobayashi, A.; Olsen, A.O.; Carrano, A.V.
submitted to the EMBL Data Library, February 1998
A:Description: Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine protease gene cluster.
A:Reference number: 214157
A:Accession: T00493
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-168 <LAM>
A:Cross-references: EMBL:AC004221; NID:g2911257; PIDN:AC004304.1; PID:g2911258
C:Genetics:
A:Gene: ATP5D
A:Map position: 19
A:Introns: 47/3; 99/1; 128/3
C:Complex: the ATP synthase F1 complex consists of three alpha chains (see PIR:FMH0D), three beta chains (see PIR:A33370), one gamma chain (see PIR:A49108), one delta chain, and one epsilon chain; the F1 complex binds to the intrinsic membrane F0 complex
C:Function:
A:Description: catalyzes the formation of ATP from ADP and phosphate using the free energy derived from proton transport down the gradient maintained by cytochrome-c oxidase across the mitochondrial inner-membrane
A:Pathway: oxidative phosphorylation
C:Superfamily: H⁺-transporting ATP synthase epsilon chain
C:Keywords: hydrolase; mitochondrion
S22348 Length: 168 February 11, 2000 15:51 Type: P Check: 1686 ..
1 MPEALLRRP GLGRVRRAR AYAEAAAPA AASGQNSF TFASPROVFF
51 NGANVAQVDV PTLGAGGIL AAHVPLQYL RGLGVVNAE DGTSTYFVS
101 SSGIANNADS SVQLAEAV TLDMLDGA KANLEKQAE LVGTADENR
151 AEQIRIEN EALYKALE
11AA_SEQUENCE 1.0
P1:D7121 - ATP synthase chain D - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 26-Aug-1999
C:Accession: D7121
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammell, C.; Fan, J.; Olinger, L.; Grimwood, J.; Davis, R.W.; Stephens, R.S.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D7121
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <ARN>
A:Cross-references: GB:AE001594; GB:AE001363; NID:g4376341; PID:g4376349
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: atpD
C:Superfamily: H⁺-transporting ATPase chain D
D7121 Length: 209 February 11, 2000 15:51 Type: P Check: 2102 ..
1 MSYQVKLRN SFLEKOLA RLOTYVPTK LKALLQAEV QNANVDAEC
51 DKDYQVAYR IYAEAEFSI PLCTDVEKS FEIOSINDF ENTAGVEVPI
101 VREVTLFPAS YSLGTPIWL DTMLSASKEL VYKKMAEVS KERLKIIEE
151 LRAVSIRVNL FEKKLIPETI KIKKIAVPL SDRSITDVGQ VKAKKRIEL
201 RKARGDECV

```

11AA_SEQUENCE 1.0
P1:531498 - ribulose-bisphosphate carboxylase (RC 4.1.1.39) small chain - potato
C:Species: Solanum tuberosum (potato)
C>Date: 03-Mar-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S31498
R:Fillz, C.C.; Wolter, F.P.; Schenkemeyer, V.; Herget, T.; Schreier, P.H.
submitted to the EMBL Data Library, December 1992
A:Description: The gene-family of ribulose-(1,5) bisphosphate
carboxylase/oxygenase small subunit from Potato.
A:Reference number: S31494
A:Accession: S31498
A:Molecule type: DNA
A:Residues: 1-181 <R1>
A:Cross-references: EMBL:X69763; NID:g21570; PIDN:CAA9417.1; PID:g21571
C:Genetics: 60/3; 105/3
A:Interons: 60/3; 105/3
C:Superfamily: ribulose-bisphosphate carboxylase small chain
C:Keywords: carbon-carbon lyase; carboxy-lyase; chloroplast
S31498 Length: 181 February 11, 2000 15:51 Type: P Check: 4433

1 MASSIVSAA VATRSVNAQA SMVAPFGLK SAASFVYTK NNNVDITSLA
51 SNGRVRQMO VMPPIMKRY ETLSTLPDL DEQLKEVEY LKNGWVPC
101 EFETHEGFV RENHKSFGY DGRWTMVKL PMFGCTDNTQ VLAEOEAKK
151 AYPQAWIRII GFQNVROVOC ISFIAYKPEG Y

11AA_SEQUENCE 1.0
P1:A71295 - probable alanine racemase (alr) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-Aug-1999
C:Accession: A71295
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Hatchman, J.M.; Hickey, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.;
Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.;
Horst, K.; Roberts, K.; Matthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; NCID:98332770
A:Accession: A71295
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-357 <COL>
A:Cross-references: GB:AE001242; GB:AE000520; NID:g3322976; PID:g3322978
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0681
C:Superfamily: alanine racemase
A71295 Length: 357 February 11, 2000 15:51 Type: P Check: 255

1 MSRTARVCL PVKADAYGHS ACQVAQAALS CGVHSEVAVC VOEASQLRRA
51 GVRAPIILCS TPTAEISSL IEHRVHTVIS ERAHIALIAR ALROSADTGA
101 TCGVHVKITD GMRIGCAPD EACALVOMVC ATPGLHLEGY CTHEVSADSV
151 RAEDLOYTEM ORAHFMHCQV YIRKSGISIP LVHANSAL LCHPRAHEDM
201 VRPGILAVGY APESVHPAVR SVFLVPMELV TOVRAIKIP ACAYVSYORL
251 WRAHETHVHG ILPIGVADGV MRALSPGLQV CIGKWPVAV GACIMDOCVY
301 DLGTPLARTV GDRVTLFGPG DAGRPGQAGD VLAHSHATIP YELCALIGKR
351 VERVYIR

11AA_SEQUENCE 1.0
P1:G70553 - probable echA10 protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Jun-1999
C:Accession: G70553
R:Coile, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.;
Krogh, A.; McLean, J.; Mole, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squires, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Squaring the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; NCID:98295987
A:Accession: G70553
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-268 <COL>
A:Cross-references: GB:X95884; GB:AL123456; NID:g3261774; PIDN:CAB09030.1;
PID:ej17233; PID:g2117180
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: echA10
C:Superfamily: naphthate synthase: enoyl-CoA hydratase homology
E:35-187/Domain: enoyl-CoA hydratase homology <ECH>
G70553 Length: 268 February 11, 2000 15:51 Type: P Check: 8877

1 MSNYRIDRT IVEGLAVTLA DGVLSVTDR PESINSITRK VLAGNADLIE
51 GAATPRVKV VRUGAGRGF SEGGAISVD VMASGPTDT VAEANRYTRA
101 IYALDPYVA VVOGPTVCG VSLALADLV LASDNAFEML AHTNGLMPD
151 GGASALYQAA IGRIRAHMA LLPDRVPAE ALSNGLVASV YPADFDPAEV
201 DKLISRLLAG PALAIAKTKN AINAATLET AFLRLRELDS QALLRLTDF
251 AEGATAFOOR RTPMFTGR

11AA_SEQUENCE 1.0
P1:H69457 - ornithine cyclodeaminase (arcB) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C:Accession: H69457
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gwin, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrleides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.F.; Dougherty, B.A.;
McKenney, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeck, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeck, R.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.;
Sikes, S.M.; Sadov, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
A:Reference number: A69250; NCID:98049343
A:Accession: H69457
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-322 <KLE>
A:Cross-references: GB:AE000988; GB:AE000782; NID:g2689311; PID:g2648890;
TIGR:AF1665
C:Superfamily: ornithine cyclodeaminase
H69457 Length: 322 February 11, 2000 15:51 Type: P Check: 4102

1 METLITQEE VESLISMDFA MNAVEAFRL YALGKAOMP KYLEFEKGD

```

51 LRAMPALHMG YAGLKWNSH PGNPDRGLPT VVALMILNSP ETGEPILAVMD
 101 ATYTSLRTG AAGGIAAKYL ARKNSVFGF ICGCTQAVFQ LEALRRVFEDI
 151 GGVKADVRE KAKKRVSYC EDRGISASVQ PAEASRCDV LVTTTPSRKP
 201 VYKAEWVEG THINAGADG PKRQELDVEI LKAKIVYDD LEQAKHGGEI
 251 NVAWSGVIC VEDVHATIGE VINGLKDRE SPEETITDS TGLAIQVAV
 301 ARVYENALMS KNVSKIKF RI

!!AA:SEQUENCE 1.0
 P1:575089 - UDP-glucose 4-epimerase - *Synechocystis* sp. (strain PCC 6803)
 N:Alternate names: protein S110244
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
 R:Accession: S75089

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hikosawa, M.; Sugiyara, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsumo, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Sliempo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3, 109-136, 1996
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S75089
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-338 <K>A>
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:JL652956; PIDN:BAH17951.1; P1D:010864; P1D:91653034

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: galE
 A:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
 E:7-332/Domain: UDPglucose 4-epimerase homology <UDP>

S75089 Length: 338 February 11, 2000 15:51 Type: P Check: 8968 ..

1 MATQOITIVT GGAGYISGSHG VIALQAGFD VLIYDNLSTG HKELVOPLG
 51 ELVVGHTGDR QKLDQLPATR NIAAVMHFAA FIAVGSYOK PDIYQNNVV
 101 GTTLLEAML AAGIKRFVFS STCAVYGMPK EIPMTESHPRI DPLSPYASK
 151 RMVEQILDVF DQAYGFSKSVI FRFENSGAD PQRGRGEBHN PETHLPLAL
 201 LVALQORPL SVFGTDYDTL DQALRDYIH VCDLAIANVL GLQYLLEGE
 251 SNIFNGNGN GFSVROYLEV AKAVTGIDLP YQLCRRRGD APLVSSAK
 301 ARELIGMNFQ YPDLHTIIDH AMQWQRRRG DYKSKNKI

!!AA:SEQUENCE 1.0
 P1:H69105 - dtp-*glucose* 4,6-dehydratase - *Methanobacterium thermoautotrophicum* (strain Delta H)
 C:Species: *Methanobacterium thermoautotrophicum*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
 C:Accession: H69105
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakey, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumu, W.; Potluer, B.; Qiu, D.; Spadafora, R.; Vicshire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani, N.; Garuso, A.; Bush, D.; Saffer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shiner, G.; Goyal, A.; Pietrowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.M.
 J. Bacteriol. 179, 7153-7155, 1997

A>Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.
 H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514

A:Accession: H69105
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-336 <MTH>
 A:Cross-references: GB:AE000933; GB:AE000666; NID:92622908; PIDN:AAB86255.1; PID:92622920
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1789
 C:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
 F:4-314/Domain: UDPglucose 4-epimerase homology <UDP>

H69105 Length: 336 February 11, 2000 15:51 Type: P Check: 3451 ..

1 MEKILVTGGA GFIGSNFIRY MQEHPYHII NDLALTYCGN LENLRGVEDE
 51 PRYTVRCSI TDRKLVDCII KQVDAYINFA AESHYDSIE DPEIFRTINI
 101 LCTQTLLEAS RKHGVRFQI ISTDEVYGA EKGFTETPT LABNSPYAS
 151 KASADLVNRA YHRTYGLPVN ITRCSNNYGP YQPEKLIPL MITNALENRP
 201 LPVYDGMNV RDMVHVLDC RAVDLVLRG RVGEVYNIG NSRRNIEIV
 251 ELVRELKGD ESTIRFEVD RGHDRYARD ASKIRNELGW KPLYSPEGI
 301 RETIRMYIDN RDMWENIKSG EYLRERYRM GGRLOD

!!AA:SEQUENCE 1.0
 P1:D69290 - dtp-*glucose* 4,6-dehydratase (rfbB) homolog - *Archaeoglobus fulgidus*
 C:Species: *Archaeoglobus fulgidus*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
 C:Accession: D69290

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.V.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kierlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.T.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 A:Reference number: A69250; MUID:98049343
 A:Accession: D69290
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-332 <KLF>
 A:Cross-references: GB:AB001082; GB:AE000782; NID:92689405; PIDN:AAB90911.1; PID:92650312; TIGR:AF0374
 C:Superfamily: *Escherichia coli* UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
 F:3-314/Domain: UDPglucose 4-epimerase homology <UDP>

D69290 Length: 332 February 11, 2000 15:51 Type: P Check: 128 ..

1 MRILVTGGG FIGSNEFVRA LAHQVEILN LDALKYGNP ENLKDAEDE
 51 RRSFKIGDIN DYELSDLIK KYDAVNFEFA ESHVDRSIS PYAFIESNFL
 101 GYITILEAVR KQNEVRLVH VSDVEYGGI EKSFTEDG LMSPPSYAS
 151 KAAAMVIVG YARTYNLAS ITRCINNYP YQPEKLIPL TIIRAMNLS
 201 VPIYTGAMV RDMVYEDHC RAVEMEMG ERREYINIS GEERTNLEIV

```

251 KTLLEIMGKD ESLIYVEDR PGHDVRYSD SWKIMRELKW RPKISFEGI
301 RTVEVYLON EWMWRPLADE RILHPTPKL RM

!!AA_SEQUENCE 1.0
P1:547045 - UDPglucose 4,6-dehydratase (EC 4.2.1.46) - Neisseria gonorrhoeae
C:Species: Neisseria gonorrhoeae
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C:Accession: S47045
R:Robertson, B.D.; Froesch, M.; van Putten, J.P.
Submitted to the EMBL Data Library, April 1994
A:Description: Identification and characterisation of cryptic rhamnose
biosynthesis genes in Neisseria gonorrhoeae.
A:Reference number: S47045
A:Accession: S47045
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <ROB>
A:Cross-references: EMBL:Z32742; NID:g520895; PIDN:CAB83652.1; PID:g520896
C:Gene: rfbB
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase
homology
C:Keywords: carbon-oxygen lyase; hydro-lyase
F:3-34/Domain: UDPglucose 4-epimerase homology <UDP>
S47045 Length: 346 February 11, 2000 15:51 Type: P Check: 5086 ..

1 MOTECKNKL VYGAGFIS AVVHITQNT RDSVNLDKL TYAGNLESIT
51 DIADNPRIAF EVDICDRAE LDRVFAQYR DAVMHLAAS HYDAISAG
101 EEIFNTIVGT EDLEEARAY WQMPSEKRE AFRFHISTD EYGDHLGTD
151 DLFTETTPYA PSSPSYASKA AADHLVRAMQ RTYRPLSYV NCSNNNGPRQ
201 FPEKLIPIMI LNALSGRPL VYGDGAQIRD WFEVDHARA LYQVYEGV
251 GETYNIGHN EKTNEVNT ICALLEELAP EKPAGVARE DLITFYDPR
301 GHDAVAVDA AKIRDLGWL PLETFESGLR KIVQWYLDNK TRONA

!!AA_SEQUENCE 1.0
P1:E70566 - Probable DRDP-GLUCOSE 4 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Jul-1999
C:Accession: E70566
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Teale, F.; Badcock, K.;
Bastam, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornby, T.; Jagers, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squires, S.; 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: E70566
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-331 <COL>
A:Cross-references: GB:Z95390; GB:AL123456; NID:g3261766; PIDN:CAB08730.1;
PID:g316065; PID:g2104387
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: rmlB
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase
homology
F:3-315/Domain: UDPglucose 4-epimerase homology <UDP>
E70566 Length: 331 February 11, 2000 15:51 Type: P Check: 5640 ..

1 MRLVGTGAG FIGTNFVSHA VREHDDAVY VIDALTYAGR RESIADVEDA
51 IRLVGDITD AELVASOLVA SDAVHPFAE SHVDNALNP EPLHNTVIG
101 TFLLEAVR HGVRLHIST DEVYGDLELD DARETESTP YNPSPYSAT
151 KAGADMVRA WRSYGVAT ISONSNNYGP YQHVKEFIPR QINVLTRGR
201 PKLGAQANY RMIVHDDN SAVRIIDRG RIGRIRYLLS EGEDNITVL
251 RFLRLMDRD PDDEFVHYDR VGHDLRAVD PSTLYDELQW APKHIDDEEG
301 LRTIDWYRD NESWRPLKD ATEARIQERG Q

!!AA_SEQUENCE 1.0
P1:G70415 - nucleotide sugar epimerase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 16-Jul-1999
C:Accession: G70415
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham,
D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman,
R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: G70415
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-321 <QOF>
A:Cross-references: GB:AE000735; NID:g2983749; PIDN:AC07310.1; PID:g2983750;
GB:AE000657
A:Experimental source: strain VFS
C:Genetics:
A:Gene: nse
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase
homology
F:3-316/Domain: UDPglucose 4-epimerase homology <UDP>
G70415 Length: 321 February 11, 2000 15:51 Type: P Check: 2675 ..

1 MNILVTCAG LIGKVKSEKL LEHGRVIGV DNINSYDPR LKEVRLBOLK
51 KFEKFEYKV NIENREALRI LFQEFEDAV INEABAGVR YSIONPHIV
101 TTNLTGNLNL LELMKEFGVK KLILASTSSL YAGQPMPEKE ELVYNTPISP
151 YVASKRAEV TAYTHYLYG IDVYLRYFT VYGLAGRPDM AVENFIYKTL
201 KQIPKXYGD GSQKRDFTY DVVAEATYKA LNKGYEIIIN VGNNKPRALK
251 ELIELIKYT GKQVKEVYGD FHKADMRDTW ADITKAKRLL GWEPKTSIEE
301 GVKRTVEWFL ENMDWVKDLR V

!!AA_SEQUENCE 1.0
P1:E6750 - glucosamine-6-phosphate isomerase homology ybft - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
C:Accession: E6750
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,
V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.;
Bourlier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brullier, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,
J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denlioz, F.; Devigne, K.M.;
Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fiebric, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizet,
A.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Goldthly, E.J.; Grandt,
G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Hensat, A.;
Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;

```

Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapins, A.; Lardinois, S.
 A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Onaga, B.; Park, S.H.; Parro, V.; Polk, T.M.; Portetelle, D.; Porcili, S.; Prescott, R.M.; Priescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Reij, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, I.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.-J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wandurt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: E69750
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-249 <KUN>
 A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12030.1; PID:el182188; PID:g2632522
 A:Experimental source: strain 168
 A:Genetics: Y
 A:Gene: ybt
 C:Superfamily: glucosamine-6-phosphate isomerase
 E69750 Length: 249 February 11, 2000 15:51 Type: P Check: 3388 ..
 1 MKIIAEHYE ELCKLSAII KEOIOAKDA VLGATGSP VGIYKOLISD
 51 YQAGEIDFSK VTFNLEDEYA GLSPSHPOSY NHEHEHLFQ HINMPDH1H
 101 IPOGDNPOLE AACKYEDLI RQAGIDVOI LGIGANGHIG FNEPESDFED
 151 RTRVAKLES TIQANARFEG GDPVLVPLRA ISMGIKTIME FSKHYILLAS
 201 GEEKADAIOR MEGPVTTOV PASILOKKNH VTVIADYKAA OKLKASFS
 I:AA_SEQUENCE 1.0
 P1:D69094 - phosphoribosylaminoimidazolesuccinocarboxamide synthase -
 Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
 A:Accession: D69094
 R:Smith, D.R.; Doucette-Stamm, L.A.; DeLoughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Glibert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pochler, B.; Qiu, D.; Spadatore, R.; Vicalre, R.; Wang, Y.; Wleczkowski, J.; Gibson, R.; Jivani, N.; Caruso, A.; Bush, D.; Sater, H.; Patwell, D.; Prabhakar, S.; McDonnell, S.; Shiner, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.D.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.
 H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: D69094
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-248 <MTH>
 A:Cross-references: GB:AE000804; GB:AE000666; NID:g2621196; PIDN:AAB84676.1; PID:g2621212
 A:Experimental source: strain Delta H
 C:Genetics: Y
 A:Gene: mth170
 C:Superfamily: phosphoribosylaminoimidazolesuccinocarboxamide synthase
 D69094 Length: 248 February 11, 2000 15:51 Type: P Check: 5550 ..

1 MDVKIDGLY SGKAKDVLIT DDEIVAVRF RDDITADGE KKDTEMKGY
 51 YNSYIAKIE EYLEAGVPT QYLERPGC ILARKLEMIP IEVIRNIIAA
 101 GSIVRRPEFT EGGEFVPLI QMDYKSDENG DPMNDIIL AGIATREBL
 151 EIKRIRLHI NSVLROFLKS RGLIIPDKL EFGEDSGRI RGDVEVSDT
 201 CRUDMEGGE PLDKDFRRG EGVGVAVRR VAKIIDDED IERNVVEL
 I:AA_SEQUENCE 1.0
 P1:A19940 - antithrombin III - baboon (fragment)
 C:Species: Papio sp. (baboon)
 C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 12-Apr-1996
 A:Accession: A19940
 R:Stackhouse, R.; Chandra, T.; Robson, K.J.H.; Woo, S.L.C.
 J. Biol. Chem. 258, 703-706, 1983
 A:Title: Purification of antithrombin III mRNA and cloning of its cDNA.
 A:Reference number: A19940; MUID:83108864
 A:Accession: A19940
 A:Molecule type: protein
 A:Residues: 1-40 <STRA>
 C:Function: A:Description: inhibits in blood plasma thrombin and activated coagulation factor X, either weakly alone or strongly in the presence of heparin
 C:Superfamily: antithrombin III
 A19940 Length: 40 February 11, 2000 15:51 Type: P Check: 1681 ..
 1 DDXNIFLSP LSVSTAFAMT KLGACNDTLK QLMKEVEKFDI
 I:AA_SEQUENCE 1.0
 P1:JX0206 - chymotrypsin inhibitor (Kunitz) WCI-3 precursor - winged bean
 C:Species: Psophocarpus tetragonolobus (winged bean)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
 A:Accession: JX0206; JN0110; JX0049; S42564; S42563
 R:Habu, Y.; Psachoknagul, S.; Umemoto, K.; Sakata, Y.; Ohno, T.
 J. Biochem. 111, 249-258, 1992
 A:Title: Structure and regulated expression of Kunitz chymotrypsin inhibitor genes in winged bean (*Psophocarpus tetragonolobus* (L.) DC.).
 A:Reference number: JX0206; MUID:92234991
 A:Accession: JX0206
 A:Molecule type: DNA
 A:Residues: 1-207 <HAB>
 R:Psachoknagul, S.; Matsui, T.; Shibata, H.; Hara, S.; Ikenaka, T.; Okada, Y.; Ohno, T.
 Plant Mol. Biol. 12, 51-58, 1989
 A:Title: Sequence and expression of the mRNA encoding the chymotrypsin inhibitor in winged bean (*Psophocarpus tetragonolobus* (L.) DC.).
 A:Reference number: JN0110
 A:Accession: JN0110
 A:Molecule type: mRNA
 A:Residues: 1-207 <PEY>
 A:Experimental source: immature seed
 R:Shibata, H.; Hara, S.; Ikenaka, T.
 J. Biochem. 104, 537-543, 1988
 A:Title: Amino acid sequence of winged bean (*Psophocarpus tetragonolobus* (L.) DC.) chymotrypsin inhibitor, WCI-3.
 A:Reference number: JX0049; MUID:89197853
 A:Accession: JX0049
 A:Molecule type: protein
 A:Residues: 25-207 <SHI>
 R:Habu, Y.; Sakata, Y.; Fukasawa, K.; Ohno, T.
 Plant Mol. Biol. 23, 1139-1150, 1993
 A:Title: Ubiquitous nuclear proteins bind to 5' upstream region of major Kunitz chymotrypsin inhibitor gene in winged bean.
 A:Reference number: S42563; MUID:94122379
 A:Accession: S42564
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-207 <HA2>
 A:Cross-references: EMBL:DJ13976; NID:g218008; PIDN:BA03086.1; PID:d1003592;

PID:g218009: EMBL:D13975; NID:g218006; PID:d1003591; PID:g218007
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, ./.
 C:Superfamily: trypsin inhibitor (Kunitz)
 C:Keywords: serine proteinase inhibitor
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-207/Product: chymotrypsin inhibitor (Kunitz) WCI-3 #status experimental
 <MAT>
 F:89/Inhibitory site: Leu (chymotrypsin) #status predicted
 JX0206 Length: 207 February 11, 2000 15:51 Type: P Check: 7697 ..
 1 MKSTFFLALF LLSAIIISLP SSTADDDLVD AEGNLVENG TYLLPHIWA
 51 HGGGIETAKT GNEPCPLTVV RSPNEVSKGE PIRISSOFLS LFIPRGSIVA
 101 LGFANPSCA ASPMTWTVDS POGPAVKLSQ OKLPENDIIV FKEFKVSHN
 151 IHVKKLYCQ HDEEDVKCQD YIGIHRDNG NRRLVTEEN PLELVLLKAK
 201 SETASSH
 !!AA_SEQUENCE 1.0
 P1:J50650 - chymotrypsin inhibitor (Kunitz) WCI-2 precursor - winged bean
 C:Species: Psophocarpus tetragonolobus (winged bean)
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 07-Feb-1997
 C:Accession: J50650
 R:Haru, Y.; Poyachoknagul, S.; Umemoto, K.; Sakata, Y.; Ohno, T.
 J. Biochem. 121, 111-249-258, 1992
 A:Title: Structure and regulated expression of Kunitz chymotrypsin inhibitor
 genes in winged bean (Psophocarpus tetragonolobus (L.) DC.).
 A:Reference number: JX0206; NUID:92244991
 A:Accession: J50650
 A:Molecule type: DNA
 A:Residues: 1-207 <HAB>
 C:Superfamily: trypsin inhibitor (Kunitz)
 C:Keywords: serine proteinase inhibitor
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-207/Product: chymotrypsin inhibitor (Kunitz) #status predicted <MAT>
 F:89-90/Region: inhibitory
 J50650 Length: 207 February 11, 2000 15:51 Type: P Check: 8167 ..
 1 MKSTFFLALF LLSAIIISLP SSTADDDLVD AEGNLVENG TYLLPHIWA
 51 HGGGIETAKT GNEPCPLTVV RSPNEVSKGE PIRISSOFLS LFIPRGSIVA
 101 LGFANPSCA ASPMTWTVDS POGPAVKLSQ OKLPENDIIV FKEFKVSHN
 151 IHVKKLYCQ RDEEDVKCQD YIGIHRDNG NRRLVTEEN PLELVLLKAK
 201 SETASSH
 !!AA_SEQUENCE 1.0
 P1:JX0246 - serine proteinase inhibitor A precursor - arrowhead
 C:Species: Sagittaria sagittifolia (arrowhead)
 C:Date: 10-Jun-1993 #sequence_revision 10-Jun-1993 #text_change 16-Jul-1999
 C:Accession: JX0246; J50780
 R:Xu, W.F.; Tao, W.K.; Gong, Z.Z.; Chi, C.W.
 J. Biochem. 119, 153-158, 1993
 A:Title: cDNA and genomic structures of arrowhead proteinase inhibitors.
 A:Reference number: JX0246; MUID:93224493
 A:Accession: JX0246
 A:Molecule type: DNA
 A:Residues: 1-181 <XU1>
 A:Cross-references: DDBJ:D13819; NID:g218280; PIDN:BAA02972.1; PID:d1003478;
 PID:g218281
 A:Note: the authors translated the codon GTC for residue 166 as Ile
 A:Accession: J50780
 A:Molecule type: mRNA
 A:Residues: 1-60, 'W', '62-83', 'S', '85-151', 'D', '153-173', 'F', '175-181 <XU2>
 A:Note: 39-Arg, 142-Val, 143-Gln and 145-His were also found
 C:Comment: This protein inhibits trypsin, chymotrypsin, elastase, and
 lysylendopeptidase.
 C:Genetics:
 A:Introns: 121/2
 C:Superfamily: arrowhead proteinase inhibitor B
 C:Keywords: serine proteinase inhibitor
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-174/Product: serine proteinase inhibitor A #status predicted <MAT>
 JX0246 Length: 181 February 11, 2000 15:51 Type: P Check: 1378 ..
 1 MAASNALLI SGVLLISLAV LCHGDPVADS DDAVQNLNG GNYPLTYIQS
 51 AAIKRGGLS TLHKDCKSY VYEAPETDRG LPVGFASAT SQPMQGSR
 101 YKFSFMPVP LICDTAMSIG KSTETGVYK LMAKSCFCK IACREVSFN
 151 VNGRTLGIG GEHFTVRFQK FDALAKTAP Q
 !!AA_SEQUENCE 1.0
 P1:J5447 - serine proteinase inhibitor A precursor - arrowhead
 C:Species: Sagittaria sagittifolia (arrowhead)
 C:Date: 10-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 07-May-1999
 C:Accession: J5447
 R:Lu, M.J.; Lu, W.Y.; Chi, C.W.
 J. Biochem. 121, 991-995, 1997
 A:Title: Clarification of an uncertain intron within the cDNA sequences of
 arrowhead proteinase inhibitors A and B.
 A:Reference number: J5447; NUID:97335950
 A:Accession: J5447
 A:Molecule type: mRNA
 A:Residues: 1-210 <LUO>
 C:Superfamily: arrowhead proteinase inhibitor B
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-210/Product: serine proteinase inhibitor A #status predicted <MAT>
 J5447 Length: 210 February 11, 2000 15:51 Type: P Check: 4632 ..
 1 MAASNALLI SGVLLISLAV LCHGDPVADS DDAVQNLNG GNYPLTYIQS
 51 AAIKRGGLS TLHKDCKSY VYEAPETDRG LPVGFASAT SQPMQGLSR
 101 YKFSFMPVP LICDTAMSIG KSETNGISF QPITADGYF LNFSEFAR
 151 STEETGVYK LAKSCFCKI ACPEVGSFNV NGRLLGIG EHFTVRFQK
 201 DALAKTAPQ
 !!AA_SEQUENCE 1.0
 P1:T06517 - alpha-amylase inhibitor 1mal precursor, monomeric - wheat
 C:Species: Triticum aestivum (common wheat)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 16-Jul-1999
 C:Accession: T06517
 R:Cardonero, P.
 submitted to the EMBL Data Library, February 1998
 A:Reference number: 215731
 A:Accession: T06517
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-151 <CAR>
 A:Cross-references: EMBL:AJ223492; NID:e1251337; PIDN:CAA11410.1; PID:e1251338
 A:Experimental source: cv. Chinese spring, endosperm
 C:Genetics:
 A:Gene: 1mal
 C:Superfamily: wheat alpha-amylase inhibitor
 C:Keywords: alpha-amylase inhibitor
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-151/Product: alpha-amylase inhibitor 1mal #status predicted <MAT>
 T06517 Length: 151 February 11, 2000 15:51 Type: P Check: 7537 ..
 1 MMKTYFQGL LVPMVAITM AVEYGARSHN SGPMSCDDPA TGYVSALTG
 51 CRAMVKLCV GSQYPAVLK DCCQQLADIN NEMRCGDLS SMLRSYQEL


```
101 GVREGKEVLP GCRKEVVKLT AASVEVCVK PIPNPSGDRA GVCYMAAYPD
151 V
!!AA_SEQUENCE 1.0
P1:S16920 - alpha-amylase inhibitor - wheat
C:Species: Triticum aestivum (common wheat)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 31-Oct-1997
C:Accession: S16920
R:Caporale, C.; Carraro, L.; Nitti, G.; Poerio, E.; Pucci, P.; Buonocore, V.
Protein Seq. Data Anal. 4, 3-8, 1991
A:Title: Determination of the primary structure of an alpha-amylase inhibitor
from wheat kernel by Edman degradation and fast atom bombardment mass
spectrometry.
A:Reference number: S16920; MUID:92020810
A:Accession: S16920
A:Status: Preliminary
A:Molecule type: Protein
A:Residues: 1-121 <CAP>
C:Superfamily: wheat alpha-amylase inhibitor
S16920 Length: 121 February 11, 2000 15:51 Type: P Check: 2886 ..

1 SGPMSWCDPA TGYKVSALTG CRAMVKLOCV GSOVPEAVLR DCCOOLADIN
51 NEMCRGCDLS SMLRSYQEL GVREGKEVLP GCRKEVVKLT AASVEVCVK
101 PIPNPSGDGA GVCYMAAYPD V
!!AA_SEQUENCE 1.0
P1:A60195 - transforming protein db1 - mouse (fragment)
N:Contains: transforming protein Mcf2
C:Species: Mus musculus (house mouse)
C:Date: 03-Mar-1993 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C:Accession: A60195; A61523
R:Galland, F.; Patis, Y.; delapeyriere, O.; Birnbaum, D.
Oncogene 6, 833-839, 1991
A:Title: Restriction and complexity of Mcf2 proto-oncogene expression.
A:Reference number: A60195; MUID:91270902
A:Accession: A60195
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-278 <GAL>
A:Note: authors translated the codon GAT for residue 229 as Ser
R:Grant, S.G.; Mattei, M.G.; Galland, F.; Stephenson, D.A.; Keltz, B.T.;
Birnbaum, D.; Chapman, V.M.
Cytogenet. Cell Genet. 54, 175-181, 1990
A:Title: Localization of the mouse Mcf-2 (db1) protooncogene within a conserved
linkage group on the mouse X chromosome.
A:Reference number: A61523; MUID:91092130
A:Accession: A61523
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-6', G', 8-73 <GRA>
C:Superfamily: db1 transforming protein; CDC24 homology; plectstrin repeat
homology
C:Keywords: oncogene; transforming protein
F:13-161/Domain: CDC24 homology #status atypical <CD24>
F:119-278/Domain: plectstrin repeat homology (fragment) <PLK>
A60195 Length: 278 February 11, 2000 15:51 Type: P Check: 2543 ..

1 DQSEKLDNSL DILKNHVLNE LIOTERAYVR ELFTVLLGYR SEMDNPOKED
51 LMPPLLRKK DVLFGNMAEI YEFHNIFMS RLEDCSDAPE RVGPCFLERK
101 DDQFMVAKYC QNKPRSELIM RYSCAFQO EQQRKLKHL GDSYLLKRV
151 QRITKYOLL KGNINELGKA VLQGSFNWL GHRKGATKK DARRKPKR
201 HLFYERAVM FCKRRESGE GADRPSPYDF KHCLKMEDVG ITYHAKGNR
251 KFEIWESEKE EIYIVAPNV DVKMLMK

101 GVREGKEVLP GCRKEVVKLT AASVEVCVK PIPNPSGDRA GVCYMAAYPD
151 V
!!AA_SEQUENCE 1.0
P1:A40316 - cyclin B - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 17-Jan-1992 #sequence_revision 17-Jan-1992 #text_change 16-Jul-1999
C:Accession: A40316
R:Buono, A.; Richardson, H.; Reed, S.I.; Russell, P.
Cell 66, 149-159, 1991
A:Title: A fission yeast B-type cyclin functioning early in the cell cycle.
A:Reference number: A40316; MUID:91300545
A:Accession: A40316
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-415 <BUE>
A:Cross-references: GB:M68881; NID:9173350; PIDN:AAA35288.1; PID:9173351
C:Superfamily: cyclin
C:Keywords: cell cycle control
A40316 Length: 415 February 11, 2000 15:51 Type: P Check: 7358 ..

1 MDVSTQTRRA TYFQDENOLQ KDHIYKRSK HIKLNTGVA PRAVDNINO
51 QDEPTLIEGN NESSISSSTG DFEEDFAYO DVEIEERSI RSTPSIGDD
101 DLENREGSFD APEGILTHGK HRLPTIPEWT KEDLALSEA AARLOANPSP
151 EDIETDPSMV PDYDEIFHY MQLERKLAP PNYMSVOOE IDWVTRMLV
201 DWIYQVQIHF RLPLETFLA VNLIDRFLSI KYVSLQKVL VGLSALLIAC
251 KYEIHPPSI YNFANVOGI FTVDEIRAE RYMLMLDID ISWGPSPFL
301 RRSIRASVD HIRMLAKYL QEVTLMEIF ISAHSEFIA TAYLSQML
351 GHDMPTCHV YXGYTAROL KPCANIWEC LVDAPNHNA IYKRSNRM
401 KRVSAPFHHM VLSVI
!!AA_SEQUENCE 1.0
P1:A57234 - 11n-44 protein precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Sep-1999
C:Accession: A57234
R:Herman, M.A.; Vassiliava, L.L.; Horvitz, H.R.; Shaw, J.E.; Herman, R.K.
Cell 83, 101-110, 1995
A:Title: The Caenorhabditis elegans gene 11n-44, which controls the polarity of
certain asymmetric cell divisions, encodes a Wnt protein and acts cell
nonautonomously.
A:Reference number: A57234; MUID:96006529
A:Accession: A57234
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-348 <HER>
A:Cross-references: GB:U22179; NID:9758776; PID:9758777
C:Superfamily: Int-1 transforming protein
A57234 Length: 348 February 11, 2000 15:51 Type: P Check: 2377 ..

1 MRAAPDEFF OSTALSTFPI LCSIATNEIP TISGAPAKI VQPKPNILK
51 QGCPSDLNS RALRSIOLAC RTHPATVISA FEGVOEGION CANLRFOOW
101 DQSEAGNIM DPPLROGR ESSLIWASS ASANGAVATA CAQGMIDCA
151 CNNOGMONEV EFGGCTHGV HGITASRKL TVGAVNTLL RYKRNILKA
201 GLAIKRTLI SSCCKHVSQ SCQKTCWRK TATLEHTDY LVEYAAKAL
251 YTDSDVYKTI DLIYLEASPD VCKAKSVAGR VCAMNETHIT QGDDRLCCG
301 NGSIRREYV RYKCDCEFW CNLVCKDCI QHRWISTCNG TPKSLIF
!!AA_SEQUENCE 1.0
```

P1:JC4152 - Wnt-11 protein precursor - chicken
 C:Species: Gallus gallus (chicken)
 C>Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 26-Aug-1999
 C:Accession: JC4152
 R:Randa, N.; Ohuchi, H.; Yoshioke, H.; Noji, S.; Nohno, T.
 Biochem. Biophys. Res. Commun. 211, 123-129, 1995
 A:Title: A chicken Wnt gene, Wnt-11, is involved in dermal development.
 A:Reference number: JC4152; MUID:95298011
 A:Accession: JC4152
 A:Molecule type: mRNA
 A:Residues: 1-354 <TAB>
 A:Cross-references: DDBJ:D31901; NID:9505349; PID:dl007271; PID:9505350
 A:Experimental source: embryo
 C:Comment: This protein is a cysteine-rich protein, and it plays an important role as a signaling factor in the cell-cell interaction during embryogenesis and as an inductive signal for dermal development.
 C:Genetics:
 A:Gene: Wnt-11
 C:Superfamily: Int-1 transforming protein
 C:Keywords: glycoprotein; mesoderm
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-354/Product: Wnt-11 protein #status predicted <MAT>
 F:40,90,304/Binding site: carbohydrate (Asn) (covalent) #status predicted
 JC4152 Length: 354 February 11, 2000 15:51 Type: P Check: 4630 ..

1 MKPSPOFFLA AFLSLILQIG ICYGIKWIAL SKTPSSIALN QTOCHKOLEG
 51 LVYSQVQLCR SNLEIMQITI QAAREVIKTC RTFSDMRWN CSSIELAPNY
 101 LLDERTRE SAFVYALSA AISHTIARAC TTGDLPGCCG GPFGEPGP
 151 GYRGGGADN LNYGLMGSK FSDAPMKMKR SSGQANKLMH LHNSEVGRVY
 201 LKASLEKCK CHGVSGSCI KTCWKGLQEL RDTALDLNKK YLSATKYVHR
 251 PMGTAKLIV KDIDIRPYKE TELLYLOSSP DECMKNEKYG SHGQDDQCN
 301 KTSNGSDSCD LMCGRGYNP YMDKVERCH CKTHMCCYVT CKKCRIVER
 351 YVCK

!!AA_SEQUENCE 1.0
 P1:S34378 - wnt-11 protein - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-Aug-1999
 C:Accession: S34378
 R:Dennis, C.L.; Delaney, S.J.; Christiansen, J.H.; Walnwright, B.J.
 Submitted to the EMBL Data Library, January 1993
 A:Description: A novel member of the murine wnt gene family (wnt-11) is expressed during mouse embryogenesis.
 A:Reference number: S34378
 A:Accession: S34378
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-354 <DEN>
 A:Cross-references: EMBL:X70800; NID:g312829; PID:g312830
 C:Superfamily: Int-1 transforming protein
 S34378 Length: 354 February 11, 2000 15:51 Type: P Check: 1689 ..

1 MRARPOVEA LFLALALHNG VCYGRKWLAL SKTPPALALN QTOCHKOLEG
 51 LVSAVOLCR SNLEIMQITI HAARQAMAC RRFADMRWN CSSIELAPNY
 101 LLDERTRE SAFVYALSA TISHTIARAC TSGDLPGCCG GPFGEPGP
 151 GNRWGGADN LSYGLMGAK FSDAPMKYK TSGQANKLMR LHNSEVGRVY
 201 LRASLETCK CHGVSGSCI RTCKWGLQEL QDVAADIKTR YLSATKYVHR
 251 PMGTAKLIV KDIDIRPYVD SELVYLOSSP DFCMKNEKYG SHGQDDQCN

301 KTSNGSDSCD LMCGRGYNP YMDKVERCH CKTHMCCYVT CKKCRIVER
 351 YVCK

!!AA_SEQUENCE 1.0
 P1:JH0687 - bone morphogenetic protein 21 precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
 C:Accession: JH0687; S16244
 R:Nishimatsu, S.; Suzuki, A.; Shoda, A.; Murakami, K.; Ueno, N.
 Biochem. Biophys. Res. Commun. 186, 1487-1495, 1992
 A:Title: Genes for bone morphogenetic proteins are differentially transcribed in early amphibian embryos.
 A:Reference number: JH0687; MUID:92378616
 A:Accession: JH0687
 A:Molecule type: mRNA
 A:Residues: 1-398 <NIS>
 A:Cross-references: GB:X63425; NID:964583; PIDN:CA445019.1; PID:964584
 A:Experimental source: oocyte
 C:Superfamily: Inhibin
 C:Keywords: glycoprotein
 F:285-398/Product: bone morphogenetic protein 21 #status predicted <MAT>
 F:137,202,237,340/Binding site: carbohydrate (Asn) (covalent) #status predicted
 JH0687 Length: 398 February 11, 2000 15:51 Type: P Check: 1620 ..

1 MVAGIHILL LOFYQILISG CTGLVPECK RYSESTNS PQSOQVLDQ
 51 FELRLNWFQ LKRRPTGKN VVIPYMLD YHLNSAQLAD DQGSSEVDYH
 101 MERAASRANT VRSFHEESM EEPESSEKT IOHFFNLSS IPDELATSS
 151 ELRIREQVQ EPEKTDGSKL HRINIYIVK PAAASRGFV VLLDTRLIH
 201 HNSKWESED VEPATIRWIA HKOPNGFVV EYTHLDNDTN VPKRHRISR
 251 SLTDKGHP RIRPLVTS HDKGHALHK ROKQARHKO RKLKSCRR
 301 HPLVDFSDV GMDWIVAPP GYHAFYCHG CPEPLADHIN STNHAIVQTL
 351 VNSVNTNPK ACCVPELSA ISMLYLDENE KVVLKNYQDM VVEGCGCR

!!AA_SEQUENCE 1.0
 P1:JH0687 - bone morphogenetic protein 21 precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
 C:Accession: JH0687; S16244
 R:Nishimatsu, S.; Suzuki, A.; Shoda, A.; Murakami, K.; Ueno, N.
 Biochem. Biophys. Res. Commun. 186, 1487-1495, 1992
 A:Title: Genes for bone morphogenetic proteins are differentially transcribed in early amphibian embryos.
 A:Reference number: JH0687; MUID:92378616
 A:Accession: JH0687
 A:Molecule type: mRNA
 A:Residues: 1-398 <NIS>
 A:Cross-references: GB:X63424; NID:964585; PIDN:CA445018.1; PID:964586
 A:Experimental source: oocyte
 R:Plessow, S.; Koester, M.; Knoechel, W.
 Biochim. Biophys. Acta 1089, 280-282, 1991
 A:Title: cDNA sequence of Xenopus laevis bone morphogenetic protein 2 (BMP-2).
 A:Reference number: S16244; MUID:91274367
 A:Accession: S16244
 A:Molecule type: mRNA
 A:Residues: 1-6, 'S', '8-15', 'V', '17-232', 'N', '234-398' <PLE>
 A:Cross-references: EMBL:X55031; NID:964581; PIDN:CA438850.1; PID:964582
 C:Superfamily: dimer; glycoprotein
 C:Keywords: dimer; glycoprotein
 F:285-398/Product: bone morphogenetic protein 21 #status predicted <MAT>
 F:137,202,340/Binding site: carbohydrate (Asn) (covalent) #status predicted
 JH0687 Length: 398 February 11, 2000 15:51 Type: P Check: 9563 ..

1 MVAGIHILL LRFQILISG CTGLVPECK RKTESGRSS PQSOQVLDQ

```

51 FELRLISNFG LKRRPPGCKN VVIPYMDL YHLHQAOLAA DGETSAMDFQ
101 MERAASRANT VRSFHHEESM EEIPESREKT IQREFFNLSS IPNEELVISA
151 ELRIFREVOQ EPFSDSSKL HRINTYDIK PAAASRGPV VRLDTRLVH
201 HNESKWESEFD VTPAIAFWIA HKOPNHFVY EVTHLDNDKN VPKKHVIRIS
251 SLTPKQDWP QIRPLVTFIS HDKGSHALHK ROKRQARHKQ RKRLKSSCRH
301 HPLVDFSDV GWNMDIYAPR GYHAFYCHGE CFPPLADHLN STNHAIVQTL
351 VNSVNTNIPK ACCVPELSA ISMLYLDENE KVLKNTQDM VVEGCCGR

!!AA_SEQUENCE 1.0
P1:S21299 protein DAF8 - eastern green mamba
C:Species: Dendroaspis angusticeps (eastern green mamba)
C:Date: 07-Dec-1994 #sequence_revision 12-May-1995 #text_change 16-Jul-1999
C:Accession: S21299
R:Duplicate: F
Submitted to the EMBL Data Library, June 1990
A:Reference number: S21298
A:Accession: S21299
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-81 <DUC>
A:Cross-references: EMBL:X53409; NID:g62755; PIDN:CAA37485.1; PID:g62756
C:Superfamily: snake toxin

S21299 Length: 81 February 11, 2000 15:51 Type: P Check: 8703

1 MKTLLTLVLT VTIYCDLAS TWICYSHKRP QPSATICEE KTCYKSVAR
51 LPAIVAGRCG GCPSKKMLVA IHCCSDKCN E

!!AA_SEQUENCE 1.0
P1:A37045 thermostable direct hemolysin homolog TRH precursor - Vibrio
parahaemolyticus
C:Species: Vibrio parahaemolyticus
C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 30-Sep-1993
C:Accession: A37045
R:Nishibuchi, M.; Taniguchi, T.; Misawa, T.; Khaemaneelam, V.; Honda, T.;
Miwatani, T.
Infect. Immun. 57, 2691-2697, 1989
A:Title: Cloning and nucleotide sequence of the gene (trh) encoding the
hemolysin related to the thermostable direct hemolysin of Vibrio
parahaemolyticus.
A:Reference number: A37045; MUID:89339712
A:Accession: A37045
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-189 <NIS>
C:Superfamily: thermostable direct hemolysin

A37045 Length: 189 February 11, 2000 15:51 Type: P Check: 6756

1 MKLKIYFAPS LLASITSVS KSFALDPSI PFPSPGSEL LFVVRNTTIK
51 TESPVNAIVN DYWTNRNIKR KPYKSVHGS IFTTSGSKWL SAYITVING
101 NNYTMAALSG YKDGSLTVFT KSEKTSLNOD FYSVKSFVDD SESSIPSINY
151 LDETPEYFVN VEAYESGNH MFWMCISNKL SFDECMSON

!!AA_SEQUENCE 1.0
P1:A48942 - thermostable direct hemolysin-related hemolysin TRH2 - Vibrio
parahaemolyticus
C:Species: Vibrio parahaemolyticus
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999
C:Accession: A48942
R:Kishishita, M.; Matsuka, N.; Kumagai, K.; Yamasaki, S.; Takeda, Y.;
Nishibuchi, M.

```

```

Appl. Environ. Microbiol. 58, 2449-2457, 1992
A:Title: Sequence variation in the thermostable direct hemolysin-related
hemolysin (trh) gene of Vibrio parahaemolyticus.
A:Reference number: A48942; MUID:92384539
A:Accession: A48942
A:Contents: AT4
A:Accession: A48942
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KIS>
A:Cross-references: GB:M88112; NID:g15305; PIDN:AAA27578.1; PID:g15306
A:Note: sequence extracted from NCBI backbone (NCBIN:112089, NCBIPI:112090)
C:Superfamily: thermostable direct hemolysin

A48942 Length: 189 February 11, 2000 15:51 Type: P Check: 5895

1 MKLKIYFAPS LLASITSVS KSFALDPSI PFPSPGSEL LFVVRNTTIK
51 TESPVNAIVN DYWTNRNIKR KPYKSVHGS VFTTASGSKL SAYITVING
101 NNYTMAALSG YKDGSLTVFT KSEKTSLNOD FYSVKSFVDD SESSIPSINY
151 LDETPEYFVN VEAYESGNH MFWMCISNKL SFDECMSON

!!AA_SEQUENCE 1.0
P1:J00472 - T-cell receptor beta chain (BRB4) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-May-1997
C:Accession: J00472
R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.
Submitted to JIPID, May 1990
A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.
A:Reference number: J00472
A:Accession: J00472
A:Molecule type: mRNA
A:Residues: 1-135 <TAN>
A:Experimental source: T cell
C:Genetics:
A:Gene: BRB4
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: receptor

J00472 Length: 135 February 11, 2000 15:51 Type: P Check: 2831

1 MISCHRODSL GAGKQALGL ALSSISPPA VSKSGASVTI ECRALDFQAT
51 TVFWYRQFPK RGLMWTASN VGSATYEOG YNKDKIPISO PDLRFSLTIV
101 TRVDPADSL YFCGVHTASH PGKMHFGPGT RLIVL

!!AA_SEQUENCE 1.0
P1:J00473 - T-cell receptor beta chain (BRB1) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 30-May-1997
C:Accession: J00473
R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.
Submitted to JIPID, May 1990
A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.
A:Reference number: J00473
A:Accession: J00473
A:Molecule type: mRNA
A:Residues: 1-136 <TAN>
A:Experimental source: T cell
C:Genetics:
A:Gene: BRB1
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: T-cell receptor

J00473 Length: 136 February 11, 2000 15:51 Type: P Check: 3911

1 MISHRODSL GATRKALGLA LSSISPPAV SKSGASVTI CRALDFQASS
51 MFWYRQFPK RGLMWTASN GTDATYEOG YNKDKIPISO DRTFSLMVI

```

101 SVDPDSSLY FCSARDSVAA DTQPLYFGAG SKLTIVL

!!AA:SEQUENCE 1.0
 P1:B53250 - class I histocompatibility antigen BOLA-AW10 alpha chain precursor
 - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 23-Jul-1999
 C:Accession: B53250
 R:Bersaud, A.; Kaushal, A.; Baldwin, C.L.; Clevers, H.; Young, J.R.; Kemp, S.J.; Machuga, N.D.; Toye, P.G.; Teale, A.J.
 Immunogenetics 33: 247-254, 1991
 A:Title: Identification of expressed bovine class I MHC genes at two loci and demonstration of physical linkage
 A:Reference number: A53250; MUID:91224708
 A:Accession: B53250
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-353 <BEN>
 A:Cross-references: GB:M69206; NID:q16345; PIDN:AA30638.1; PID:q289420
 C:Superfamily: class I histocompatibility antigen; Immunoglobulin homology
 F:210-275/Domain: Immunoglobulin homology <IMM>

B53250 Length: 353 February 11, 2000 15:51 Type: P Check: 6952 ..

1 LLLSGVLVT ETRAGSHSMR YFTAVSRPG LGEPRTLEVG YVDDTQVFR

51 DSDAEPNPMR PRARWEOEG PEYMDRETQR AKGNAOFFRV SLNLRGYN

101 OSEAGSHTLQ WMSGCYVGPD GRPRGFMOF GYDGRDYLAL NEDLRSMTAV

151 ETMAQISRRK MEAGAEAVQ RNYLGRVCE WLRRYLENGR DTLRADPPK

201 AHVTRHPISG REYTLRCMAL GFYREISLT WORGEDQDQ DMELVETRPS

251 GDNQFQWAA LTVSGEEQK YTCQVHEGL QEPPLTKWEP POPSELTWGI

301 IYGLVLAVT GAVYAGVIC MKRSRGERG TYIASSSDS AGSDVSLTV

351 PKV

!!AA:SEQUENCE 1.0
 P1:15609 - T-cell surface glycoprotein CD8 beta chain - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 23-Jul-1999
 C:Accession: 150609; S37237
 R:Teegases, C.A.; Kong, F.K.; Paramithiotis, E.; Chen, C.L.; Ratcliffe, M.J.; Davison, T.F.; Young, J.R.
 J. Immunol. 154: 4485-4494, 1995
 A:Title: Identification and analysis of the expression of CD8 alpha beta and CD8 alpha alpha isoforms in chickens reveals a major TCR-gamma delta CD8 alpha beta subset of intestinal intraepithelial lymphocytes.
 A:Reference number: 150609; MUID:95238946
 A:Accession: 150609
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-207 <TR>
 A:Cross-references: EMBL:Z26484; NID:9403300; PIDN:CA81258.1; PID:9403301
 C:Superfamily: T-cell surface glycoprotein CD8 beta chain; Immunoglobulin homology
 C:Keywords: glycoprotein

150609 Length: 207 February 11, 2000 15:51 Type: P Check: 8972 ..

1 MARPMWML CLQLPFCIN LLSQTPGYI LTKTNNSPEI VCPMKSEHNG

51 VVYRNNGR QHFEFLFS PLGKATYGN ISOEKFSIRG TSSYHSYRLH

101 INRLHSDNG TYCCITIOS QLLIGTQL DVEDLPLPS MSLLVPLTKK

151 PMCRKXKNA INKKGACTPM VWPPLAAGAL LLLSLIPTI RRRYRLRRRL

201 WYFAHR

!!AA:SEQUENCE 1.0
 P1:139516 - ribosomal protein S3 - Acholeplasma axanthum
 C:Species: Acholeplasma axanthum
 C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 13-Aug-1999
 C:Accession: 139516
 R:Roth, K.T.; Harrison, N.; Sears, B.B.; Int. J. Syst. Bacteriol. 44: 119-124, 1994
 A:Title: Phylogenetic relationships among members of the class Mollicutes deduced from rps3 gene sequences.
 A:Reference number: 139515; MUID:94169035
 A:Accession: 139516
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-257 <RES>
 A:Cross-references: GB:L22465; NID:9456689; PIDN:AAA21916.1; PID:9456690
 C:Genetics:
 A:Gene: rps3
 C:Superfamily: Escherichia coli ribosomal protein S3

139516 Length: 257 February 11, 2000 15:51 Type: P Check: 7310 ..

1 MGRKNPPIGF RVGVRDMS KMYADKIVP ALVKEDAVIR KPLNKNYNA

51 AVSHIEIRL KELKYKRVR ILHSGKPCV VIGFEALIMK ETIASLEKIT

101 KKEIVFNVE VRKEPVATL VAQSMADOLE NRASFERSK NCYAKSIKVR

151 AKGITLSOR LGREMARKTE GYSEGOVPLH TLRAVETAT ARAQTTGIL

201 GIKWVIHGE ILPGQSHHEL KREROSSASS NHGGKRRPS RKGPRRQED

251 AATEGN

!!AA:SEQUENCE 1.0
 P1:S77489 - ribosomal protein S8 - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein s11809
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 13-Aug-1999
 C:Accession: S77489
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Aamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3: 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S77489
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Residues: 1-113 <RAN>
 A:Molecule type: DNA
 A:Cross-references: EMBL:D90905; GB:AB001339; NID:q1652360; PIDN:BA17336.1; PID:d1018059; PID:q1652414
 A>Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: rps8
 C:Superfamily: Escherichia coli ribosomal protein S8
 C:Keywords: protein biosynthesis; ribosome

S77489 Length: 133 February 11, 2000 15:51 Type: P Check: 8926 ..

1 MASTDIDSM LTRIRNACV RHSTQVPTT KWTLSIAKVL KSEGFEDYS

51 ETGEGINKML VTLTKYKGT RQPLINTLOR VSKPOLRVYS PSKKIPRVLS

101 GIGIAIVTS HGIMTDREAR ROGIGGEIIC YIW

!!AA:SEQUENCE 1.0
 P1:701947 - hypothetical protein F104.1 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 13-Aug-1999
 C:Accession: T01947
 R:Abu-Threideh, J.; Stoneking, T.; Langston, Y.; Trevaakis, E.
 Submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of A. thaliana F1104.
 A:Reference number: 214466
 A:Accession: T01947
 A:Status: Preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-255 <ABD>
 A:Cross-references: EMBL:AF096370; NID:g3695372; PIDN:AAC62776.1; PID:g3695375
 C:Genetics:
 A:Map position: 4
 A:Note: F1104.1
 C:Superfamily: Lens fiber membrane major intrinsic protein

T01947 Length: 252 February 11, 2000 15:51 Type: P Check: 4063 ..
 1 MPINRIAIGT PGEASRPDAI RAFAEFPSM VIFVPAQGS GMAYGKLTGD
 51 GPATPAGLVA ASLSHAFALF VAASVGANVS GGHVNDVATF GAFIGNTTL
 101 LRAILWIAQ LGAVVACLL LKVGSGMET AAFSLSYGT PMNAVFEIV
 151 MFGVLYTVV ATAVDPKRGD IGIAPLAIG LIVGANILVG GAFDGSNMP
 201 AVSFGPAVVS WIMTNHMYWV VGPFGAIA AIYVDIIFIG SNGHEPLPSN
 251 DF

11AA_SEQUENCE 1.0
 P1:A39234 - opsin - bluebottle fly (Calliphora vicina)
 C:Species: Calliphora vicina
 C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 13-Aug-1999
 C:Accession: A39234
 R:Huber, A.; Smith, D.P.; Zuker, C.S.; Paulsen, R.
 J. Biol. Chem. 265, 17906-17910, 1990
 A:Title: Opsin of Calliphora peripheral photoreceptors R1-6. Homology with Drosophila Rh1 and posttranslational processing.
 A:Reference number: A39234; MUID:91009258
 A:Accession: A39234
 A:Molecule type: mRNA
 A:Residues: 1-371 <HUB>
 A:Cross-references: GB:M58334; GB:J05596; NID:g156511; PIDN:AAA62725.1; PID:g156512
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: chromoprotein; G protein-coupled receptor; glycoprotein; lipoprotein; membrane protein; photoreceptor; retinal
 F:317/Binding site: retinal (Lys) (covalent) #status predicted
 A39234 Length: 371 February 11, 2000 15:51 Type: P Check: 9791 ..

1 MERKSTFLIG PSFALITNGS VTDKVTDPMA HLHVHPWNOF PAMEKMAKAF
 51 LAAYMTIAT ISMCGNGVY YIFSTISLR TPANLVLINL AISDEGIMIT
 101 NTPMAGINLF YETWVLGPLM CDYIGGLGSA FCGSSILSMC MISDRNVT
 151 VKMGAGPMT IKLAIKIAL IWMASITWL APVFGSKRYV PEGNITSCGI
 201 DYLERDWNPR SYLIFYSTIFV YLPLFLICY SYWITIAVS AHEKAMRQA
 251 KKNVNSLRS SEDADKSAEG KLAQVALVTI SLWFAMWTFY TIINTLGLFK
 301 YEGITPLNTI WGACFAKSA CYNPIYVIGS HPRYGIALKE KCPCCVEGKY
 351 DDGKASDATS QATNNESEK A
 11AA_SEQUENCE 1.0
 P1:S40691 - opsin rh1 - fruit fly (Drosophila pseudoobscura)
 C:Species: Drosophila pseudoobscura
 C:Date: 18-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
 C:Accession: S40691; S65546; S24605

R:Carulli, J.P.; Hartl, D.L.
 Genetics 132, 193-204, 1992
 A:Title: Variable rates of evolution among Drosophila opsin genes.
 A:Reference number: S40691; MUID:93012921
 A:Accession: S40691
 A:Molecule type: DNA
 A:Residues: 1-374 <CAR>
 A:Cross-references: EMBL:X65877
 R:Carulli, J.P.
 Submitted to the EMBL Data Library, November 1992
 A:Accession number: S65546
 A:Accession: S65546
 A:Molecule type: DNA
 A:Residues: 1-173 'W', 175-374 <CAR>
 A:Cross-references: EMBL:X65877; NID:99077; PIDN:CAA46708.1; PID:99078
 C:Genetics:
 A:Gene: FlyBase:Dpse/nhae
 A:Cross-references: FlyBase:FBgn0012733
 A:Introns: 3/2; 191/2; 240/3; 333/2
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: chromoprotein; lipoprotein; retinal
 F:320/Binding site: retinal (Lys) (covalent) #status predicted
 S40691 Length: 374 February 11, 2000 15:51 Type: P Check: 7181 ..

1 MDSFAAVATQ LGQFAFASN GSVVDKVTDP MAHLISPYWD QPAMPPIWA
 51 KILTAYMIIT GMISWCGNGV VIYIFATRS LITPANLVI NLAISDEGIM
 101 ITNTPMAGIN LYFETWVLGP MMCDIYAGLG SAFGCSISMS MCMISLDRYQ
 151 VIYKMGARBP MTIPLALCKI AYICFMSTIM CCLAVFGMS RYVDEGNLTS
 201 CGIDYLEROM NPSRLYIFYS IFYIYIPLF ICYSWIFIA AVASHEKAMR
 251 EQAKKNVKS LRSSBDADKS ABEKLAQVAL VTISLWFAMW TPLIVINCNG
 301 LFPEDELTP LNTIMGACPAK SAACYNPIYV GISHPKRYLA LKEKPCCVF
 351 GKVDGKSSSE AOSQATTSEA ESKA
 11AA_SEQUENCE 1.0
 P1:JC4304 - Orphan G-protein-coupled receptor - human
 N:Alternate names: V28 protein
 C:Species: Homo sapiens (man)
 C:Date: 16-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 20-Sep-1999
 C:Accession: JC4304
 R:Rapport, C.J.; Schweickart, V.L.; Eddy Jr., R.L.; Shows, T.B.; Gray, P.W.
 Gene 163, 295-299, 1995
 A:Title: The orphan G-protein-coupled receptor-encoding gene V28 is closely related to genes for chemokine receptors and is expressed in lymphoid and neural tissues.
 A:Reference number: JC4304; MUID:96011651
 A:Accession: JC4304
 A:Molecule type: mRNA
 A:Residues: 1-355 <RAP>
 A:Cross-references: GB:U20350; NID:g665580; PID:g665581
 A:Experimental source: peripheral blood mononuclear cell
 C:Comment: This protein is a cell-surface receptor which recognizes extracellular signals and transduces those signals into an intracellular response.
 C:Comment: This protein is a key regulator of many immune and homeostatic responses, and interacts between the nervous and immune systems.
 C:Genetics:
 A:Gene: V28
 A:Map position: 3pter-p21
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; lymphokine; transmembrane protein
 F:35-57/Domain: transmembrane #status predicted <TM1>
 F:66-88/Domain: transmembrane #status predicted <TM2>
 F:104-125/Domain: transmembrane #status predicted <TM3>
 F:146-165/Domain: transmembrane #status predicted <TM4>
 F:197-217/Domain: transmembrane #status predicted <TM5>

Mon Feb 14 08:07:32 2000

pir.cat

Page 91

F:230-254/Domain: transmembrane #status predicted <TM6>
F:275-296/Domain: transmembrane #status predicted <TM7>

JC4304 Length: 355 February 11, 2000 15:51 Type: P Check: 2863

```
1 MQGFESTATE NFEYDGLAE CYIGDIYVFG TVFLSTIFYSV IFAIGLVNL
51 LVEVALTMSK KPKSVTDIYL LNLALSDLE VALPEWTHY LINEKGLHNA
101 MCKFTTAPFF IGFGSIEFI TVISIDRYLA IYLAANSMN RTVOHGVTIS
151 LGVMAALIV AAPQFMETK KENECLGYP EVLGEIWPVL RAVETNEFLGF
201 LLEPLIMSC YFRITOTLES CNHKKAKAI KLILVIVF FLEWPIYVM
251 ILEETIKLYD FPGCDMKRD LRALSVTEI VAFSHCLNP LIAAFAGEKF
301 RRYLHLYGK CLAVLCGRSV HDFSSESGO RSRHGVLS NRTYHSDGD
351 ALLLL

!!AA_SEQUENCE 1.0
P1:158186 - hypothetical G-protein coupled receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 26-Aug-1999
C:Accession: 158186
R:Harrison, J.K.; Barber, C.M.; Lynch, K.R.
Neurosci. Lett. 169, 85-89, 1994
A:Title: cDNA cloning of a G-protein-coupled receptor expressed in rat spinal
cord and brain related to chemokine receptors.
A:Reference number: 158186; MUID:94323113
A:Accession: 158186
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-354 <RES>
A:Cross-references: EMBL:U04808; NID:9439860; PID:9439861
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor

158186 Length: 354 February 11, 2000 15:51 Type: P Check: 4478
```

```
1 MPTFPEIDL ENFEYDSNAE ACYLGDIYAF GTFFLSIEFVS LVTFGLVGN
51 LLVVALTNS RSKSTITDIY LNLALSDLE EVATLPEFWH YLISHEGLHN
101 AMCKLTAPFF FICFGGICIF ITVISIDRYL AYLAAANSMN NRTVOHGVTI
151 SLGVMAAIL VASQFMETK RKNDECLDY PEVLQIWPV LRNESEVNLG
201 FVLPLIMSF CYFRIVRILE SCKNKKAKA IRILLVIVY FLEWPIYNI
251 VIFLETLKFY NFPPSCGMKR DIRMALSTEY TVAFSHCLN PRIVAFAGEK
301 FRRYLRHLYN KCLAVLCGRP VHAQFSTESQ RSKDSTILSS LTHYTSSEGS
351 SLLL

!!AA_SEQUENCE 1.0
P1:A29667 - pulmonary surfactant protein B - bovine (fragment)
M:Alternate names: pulmonary surfactant 1-phe protein
C:Species: Bos primigenius taurus (cattle)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 16-Oct-1998
C:Accession: A29667; S02317
R:Olafson, R.W.; Rink, U.; Ktjelland, S.; Yu, S.H.; Chung, J.; Harding, P.G.R.;
Possmayer, F.
Biochem. Biophys. Res. Commun. 148, 1406-1411, 1987
A:Title: Protein sequence analysis studies on the low molecular weight
hydrophobic proteins associated with bovine pulmonary surfactant.
A:Reference number: A29667
A:Accession: A29667
A:Molecule type: Protein
A:Residues: 1-60 <OAA>
R:Yu, S.H.; Chung, W.; Olafson, R.W.; Harding, P.G.R.; Possmayer, F.
```

Biochim. Biophys. Acta 921, 437-448, 1987
A:Title: Characterization of the small hydrophobic proteins associated with
pulmonary surfactant.
A:Reference number: S02317; MUID:88025156
A:Accession: S02317
A:Molecule type: Protein
A:Residues: 1-10 <YU>
C:Superfamily: pulmonary surfactant protein B; saposin repeat homology

A29667 Length: 60 February 11, 2000 15:51 Type: P Check: 8980

```
1 FPIPIPYCML LRLIKKIQK VIPKGVLTMT VAVGVHVPV LVGGIIQOLV
51 IEXSVILMTD

!!AA_SEQUENCE 1.0
P1:S2494 - RNA N-glycosidase (EC 3.2.2.22) alpha-luffin precursor - smooth
loofah
N:Alternate names: protein synthesis inhibitor; ribosome-inactivating protein
luffin-A
C:Species: Luffa cylindrica (smooth loofah)
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Aug-1999
C:Accession: S22494; S26390; JH0202; A32542
R:Kataoka, J.; Habuka, N.; Miyano, M.; Masuta, C.; Koiwai, A.
Plant Mol. Biol. 18, 1199-1202, 1992
A:Title: Nucleotide sequence of cDNA encoding alpha-luffin, a
ribosome-inactivating protein from Luffa cylindrica.
A:Reference number: S22494; MUID:92288316
A:Accession: S22494
A:Molecule type: mRNA
A:Residues: 1-277 <KAT>
A:Cross-references: EMBL:X62371; NID:919145; PID:CA44229.1; PID:919146
R:Islam, M.R.; Nishida, H.; Funatsu, G.
Agric. Biol. Chem. 54, 2967-2978, 1990
A:Title: Complete amino acid sequence of luffin-a, a ribosome-inactivating
protein from the seeds of sponge gourd (Luffa cylindrica).
A:Reference number: S26390; MUID:91248471
A:Accession: S26390
A:Molecule type: protein
A:Residues: 1-53, 'L', '55', 'I', '57-86', 'SOL', '90-154', 'L', '156-157', 'I', '159-173', 'L', '175-209', 'SL', '212
-267' <ISD>
R:Islam, M.R.; Nishida, H.; Funatsu, G.
Agric. Biol. Chem. 54, 1343-1345, 1990
A:Title: Complete amino acid sequence of luffin-a, a ribosome-inactivating
protein from the seeds of Luffa cylindrica.
A:Reference number: JH0202; MUID:91197482
A:Accession: JH0202
A:Molecule type: protein
A:Residues: 21-53, 'L', '55', 'I', '57-86', 'SOL', '90-154', 'L', '156-157', 'I', '159-173', 'L', '175-209', 'SL', '211
-267' <IS2>
A:Experimental source: seed
R:Ramakrishnan, S.; Enghild, J.J.; Bryant Jr., H.L.; Xu, F.J.
Biochem. Biophys. Res. Commun. 160, 509-516, 1989
A:Title: Characterization of a translation inhibitory protein from Luffa
aegyptiaca.
A:Reference number: A32542; MUID:89246493
A:Accession: A32542
A:Status: preliminary
A:Molecule type: Protein
A:Residues: 21-28, 'G', '30', 'X', '32-34', 'K', '36-40' <RAM>
C:Function: hydrolyzes the N-glycosidic bond of a specific adenosine in 28S
rRNA thereby inactivating the ribosome
A:RNA thereby inactivating the ribosome
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology
C:Keywords: glycoprotein; glycosidase; hydrolase; seed; toxin
F:1-26/Domain: signal sequence #status predicted <SIG>
F:21-267/Product: RNA N-glycosidase alpha-luffin #status experimental <MAT>
F:24-262/Domain: RNA N-glycosidase homology <RNG>
F:268-277/Domain: carboxyl terminal propeptide #status predicted <CTP>
F:48, '53', '97', '104', '225', '246/Binding site: carbohydrate (Asn) (covalent) #status
experimental
```

F:90.179.182/Active site: Tyr, Glu, Arg #status predicted

S22494 Length: 277 February 11, 2000 15:51 Type: P Check: 429 ..

1 MKRFTVLIA IFVASTVEA DVAFSLSGSS STSKSFIFD LRRALPSNGT
C:Species: ASGARSTLM TSNYDGKAI TVADVNTVY IMGLVNSTS
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Aug-1999
C:Accession: S23519, S23113
R:Kataoka, J.; Haiduka, N.; Miyano, M.; Masuta, C.; Koiwai, A.
Plant Mol. Biol. 19, 887-889, 1992
A:Title: Nucleotide sequence of cDNA encoding beta-luffin, another
ribosome-inactivating protein from *Luffa cylindrica*.
A:Reference number: S23519; MUID:92353400
A:Accession: S23519
A:Molecule type: mRNA
A:Residues: 1-278 <KAT>
A:Cross-references: EMBL:X62372; NID:919149; PIDN:CAA44230.1; PID:919150
A:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology
F:26-264/Domain: RNA N-glycosidase homology <RNG>

11AA_SEQUENCE 1.0
P1:JN0108 - luffin-b - smooth loofah
C:Species: *Luffa cylindrica* (smooth loofah)
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 07-May-1999
C:Accession: JN0108
R:Islam, M.R.; Hirayama, H.; Funatsu, G.
Agric. Biol. Chem. 55, 229-238, 1991
A:Title: Complete amino acid sequence of luffin-b, a ribosome-inactivating
protein from sponge gourd (*Luffa cylindrica*) seeds.
A:Reference number: JN0108; MUID:91248488
A:Accession: JN0108
A:Status: Preliminary
A:Molecule type: Protein
A:Residues: 1-250 <ISL>
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology
F:5-246/Domain: RNA N-glycosidase homology <RNG>

251 SKVTNINIKL LNKONIAAF DDGIPTKH
201 VSPALSL ENEWSALSKQI QLAQTNGAF RTPVITDKR GORVEIKDYN
151 RAFDASITSL FHYDSTAAG AFLVITQTTA EASREKIEG QITERPKNE
101 TSYFNEDDA KLASQYFKG STIVLPYSG NYERLQNAAG KVEKIPLEG
51 EVKSNIPILL PSASGASRYI LMGLSYDAK AITMAIDVTN VYIMGLVNS

11AA_SEQUENCE 1.0
P1:JN0108 - luffin-b - smooth loofah
C:Species: *Luffa cylindrica* (smooth loofah)
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 07-May-1999
C:Accession: JN0108
R:Islam, M.R.; Hirayama, H.; Funatsu, G.
Agric. Biol. Chem. 55, 229-238, 1991
A:Title: Complete amino acid sequence of luffin-b, a ribosome-inactivating
protein from sponge gourd (*Luffa cylindrica*) seeds.
A:Reference number: JN0108; MUID:91248488
A:Accession: JN0108
A:Status: Preliminary
A:Molecule type: Protein
A:Residues: 1-250 <ISL>
C:Superfamily: RNA N-glycosidase; RNA N-glycosidase homology
F:5-246/Domain: RNA N-glycosidase homology <RNG>

JN0108 Length: 250 February 11, 2000 15:51 Type: P Check: 7681 ..
1 ANVSHSLSGA DSKSYSEFI ALRKALPSKE KVSNIPLLP SASGASRYIL
51 MOLSNDYDKA ITMAIDVTN VYIMGLVNST SYFNESDAK LASQYFKGS
101 TLVTPYSGN YERLQNAAG IREKIPLEGF ALDSALTSIF HYDSTAAGAA
151 FLVITQTTAE ASRKYIEGQ IIERIPKNEY PSPALSL ENWSLTSKQI

201 QLAQTNGAF RTPVITDKR GORVEITNLA SKVOIKDVS KLLNKONIA

11AA_SEQUENCE 1.0
P1:S65052 - pistil-specific protein sts14 precursor - potato
C:Species: *Solanum tuberosum* (potato)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Aug-1999
C:Accession: S65052
R:van Eridik, G.J.; Wagens, M.; Ruiter, R.K.; van Herpen, M.M.A.; Schrauwen,
J.A.M.; Willems, G.J.
Plant Mol. Biol. 30, 171-176, 1996
A:Title: Molecular analysis of a pistil-specific gene expressed in the stigma
and stilar cortex of *Solanum tuberosum*.
A:Reference number: S65052; MUID:96197407
A:Accession: S65052
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-214 <VAN>
A:Cross-references: EMBL:X82652; NID:91236784; PIDN:CAA57976.1; PID:9124149;
PID:91236785
C:Gene: sts14
C:Superfamily: pathogenesis-related leaf protein

S65052 Length: 214 February 11, 2000 15:51 Type: P Check: 5285 ..

1 MEVLSTAMAC LVYIYIYD EKKRELEKVR NKMNTLFFQ FLLTTASSL
51 THISAQTVP PPPTPSAAT PPSRAOEFL DAHNKARSEV GVGPLTWSM
101 LAKETSLVR YQRDKONCF ANLSNGKYG NQLMASGIV TPRMAVDSV
151 AEKRYNEN NSCTGDDKCG VYTOYWKKS IEIGCAQRTIC YEGPATLVC
201 FYNPGNVIG EKPX

11AA_SEQUENCE 1.0
P1:JC4822 - acyl carrier protein - *Bacillus subtilis*
N:Alternate names: 8.5K protein
C:Species: *Bacillus subtilis*
C:Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 20-Aug-1999
C:Accession: JC4822; A69582
R:Oguero, A.; Kakashita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.
Gene 172, 17-24, 1996
A:Title: The effect of Srb, a homologue of the mammalian SRP receptor
alpha-subunit, on *Bacillus subtilis* growth and protein translocation.
A:Reference number: JC4819; MUID:96257247
A:Accession: JC4822
A:Molecule type: DNA
A:Residues: 1-77 <OGU>
A:Cross-references: DDBJ:D64116; NID:91389548; PIDN:BAI0975.1; PID:91011632;
PID:91337013
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,
V.; Galletto, M.G.; Besleres, P.; Bolicin, A.; Borchert, S.; Boriss, R.;
Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, A.;
J.J.; Conneron, I.F.; Cummings, N.J.; Daniel, R.A.; Denicourt, F.; Devine, K.M.;
Duesterhoft, A.; Ertlach, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Faber, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,
A.; Gallenro, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golligly, E.J.; Grand,
G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.;
Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;
Joris, B.; Karamata, D.; Kasahara, Y.; Klaert-Blanchard, M.; Klein, C.;
Kobayashi, Y.; Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurlta, K.;
Lapidus, A.; Lardinois, S.
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
S.; Maueel, C.; Medina, C.; Medina, N.; Mellado, R.P.; Mizuno, H.; Moestl, D.;
Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega,
B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Portwick, S.; Prescott,
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,
S.; Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato,

T.: Scanlon, E.
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serrier, P.; Shin, B.S.; Sojda, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpilava, P.; Tognoni, A.; Tosato, V.; Uchiyama, M.; Vandenberg, M.; Vannier, F.; Vassart, A.; Vlah, A.; Wambach, R.; Weder, E.; Weder, H.; Wetzinger, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: A69582
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-77 <RDN>
A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PIDN:CAB13465.1; PID:el185183; PID:92633964
A:Experimental source: strain 168
C:Genetics:
A:Gene: acpA; strb
C:Function:
A:Description: carrier of the growing fatty acid chain; growing fatty acid chain is covalently bound to 4'-phosphopantetheine prosthetic group
A:Pathway: fatty acid biosynthesis
C:Superfamily: acyl carrier protein; acyl carrier protein homology
C:Keywords: fatty acid biosynthesis; phosphopantetheine; phosphoprotein
F:2-73/Domain: acyl carrier protein homology <ACP>
F:37/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
Jc4822 Length: 77 February 11, 2000 15:51 Type: P Check: 8148 ..

1 MADLERYTR IYDRUGYDE ADVKLEASF EDGANSUDV VELVMELEDE
51 FDMISDEDA EKATYGDVAV NYINOC
!!AA:SEQUENCE 1.0
P1:B65003 histidine transport system permease protein HisQ - *Escherichia coli* (strain K-12)
C:Species: *Escherichia coli*
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 20-Aug-1999
C:Accession: B65003
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: B65003
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-228 <BLAT>
A:Cross-references: GB:AE000319; GB:U00096; NID:91788634; PIDN:AAC75368.1; PID:91788646; UNCP:B2308
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: hisQ
C:Superfamily: histidine permease protein M
B65003 Length: 228 February 11, 2000 15:51 Type: P Check: 7651 ..

1 MUXGSSVIL OGALVLELA ISSVVLAVII GLIGAGKUS QNRSLGLFE
51 GYTLIRGVP DVLMLIIF GLOALNVT EAMGVOIDI DMVAGITL
101 GFIVAYITE TFGKFAVAV KKHIAAFAF GTRGVQFRR IMFPMKXA
151 LFGINNMVY ILKSTALVSL LGLDEVYKAT QLAGKSTWEP FTFALVCYI
201 YLVFTVSNG VLLFLERRIS VGVKRAAL
!!AA:SEQUENCE 1.0
P1:B36263 - oligopeptide transport system permease protein oppB - *Escherichia coli*
C:Species: *Escherichia coli*
C>Date: 18-Jan-1991 #sequence_revision 30-Sep-1997 #text_change 20-Aug-1999
C:Accession: G64871; B36263
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64871
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-306 <BLAT>
A:Cross-references: GB:AE000223; GB:U00096; NID:91787496; PIDN:AAC74326.1; PID:91787497; UNCP:B1244
A:Experimental source: strain K-12, substrain MG1655
R:Kashiyagi, K.; Yamaguchi, Y.; Sakai, Y.; Kobayashi, H.; Igatahshi, K. J. Biol. Chem. 265, 8387-8391, 1990
A:Title: Identification of the polyamine-induced protein as a periplasmic oligopeptide binding protein.
A:Reference number: A36263; MUID:90256748
A:Accession: B36263
A:Molecule type: DNA
A:Residues: 1-16 <KAS>
A:Cross-references: GB:U05433
C:Genetics:
A:Gene: oppB
C:Function:
A:Description: probably responsible for mediating passage of peptides across the cytoplasmic membrane
C:Superfamily: oligopeptide permease protein oppB
C:Keywords: binding protein; protein-dependent transport system; oligopeptide transport; transmembrane protein
F:9-25/Domain: transmembrane #status predicted <TM1>
F:105-121/Domain: transmembrane #status predicted <TM2>
F:141-157/Domain: transmembrane #status predicted <TM3>
F:173-189/Domain: transmembrane #status predicted <TM4>
F:228-244/Domain: transmembrane #status predicted <TM5>
F:273-289/Domain: transmembrane #status predicted <TM6>
B36263 Length: 306 February 11, 2000 15:51 Type: P Check: 5147 ..

1 MKFLIRKCL EALPTLFIIL TISFPMKLA PGSPFGER LPEVMANIE
51 AKYHLNDPIM TQFSTYKQL AHGDFGPEFK YKDSVNDLV ASSFYSAKL
101 GAAPFLAVI LGVSAGVIAA LKONTRWDYT VMGLAMGVV IPSEVAPDL
151 VMFAIILHW LPGGWNNGA LKFMILPVA LSLAYINASIA RITRSMIEV
201 LHSNFIPTAR AKGLPMRII LRHALKPLL PVLSYMGAP VGIITGSNVI
251 ETIYGLPGIG QLFVNGALNR DYSLVSLTI LVGALTILFN AIVDLYAVI
301 DPRIRY
!!AA:SEQUENCE 1.0
P1:C71163 - probable oligopeptide transport system permease protein AppC - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 26-Aug-1999
C:Accession: C71163
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamaya, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3.
A:Reference number: A71000; MUID:98344137

A:Accession: C71163
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Reference type: DNA
 A:Residues: 1-304 <RAW>
 A:Cross-references: GB:AP000002; NID:g3236129; PID:d1030535; PID:g3256909
 A:Experimental source: strain O73
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Gene: PHO504
 C:Superfamily: oligopeptide permease protein oppb

C71163 Length: 304 February 11, 2000 15:51 Type: P Check: 4618

1 MEVTKLAFNR NKKFGFGML IAFITFGIG PLFTFPASDG LYEEYQKIK
 51 IASYSAKTLP PMTRENITTY TGRKEVLIHI LGTDKCKDL YTELYIGLRT
 101 SLWVAFLAAI IGTITGITIG FVSGYKGLV DELIMFVNI MLVPSIVLL
 151 ILVAAYLEAR SPEIGALITG LTNMPVARS VRAQTLSLN REVMKSKIM
 201 GVGDRLRIIE DILPNMISYI FMTGILQVSG AILASATLDF IGLGPTTMS
 251 LGVILQAIM HNALQFGWMW WFIPPGLFIT LIITSLFFIN LGLEEVNPR
 301 LRGV

11A:SEQUENCE 1.0
 P1:S64728 - protein secretion protein xcp - Pseudomonas putida
 C:Species: Pseudomonas putida
 C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 20-Aug-1999
 C:Accession: S64728; S47505
 R:de Grooc, A.; Kijner, J.J.; Filloux, A.; Tommassen, J.
 Mol. Gen. Genet. 250, 491-504, 1996
 A:Title: Characterization of type II protein secretion (xcp) genes in the plant growth-stimulating Pseudomonas putida, strain WCS358.
 A:Reference number: S64724; MUID:96186881
 A:Accession: S64728
 A:Molecule type: DNA
 A:Residues: 1-482 <DEG>
 A:Cross-references: EMBL:X81085; NID:g531737; PID:CAAS6980.1; PID:g531741
 C:Genetics:
 A:Gene: xcpR
 C:Superfamily: secretion protein xcpR

S64728 Length: 482 February 11, 2000 15:51 Type: P Check: 6546

1 MLPYRLARQA GLAMPAGQG WQMLRRDAD SEVLQELNRY HGQPSALAYL
 51 EPAAFDQQLA QLYGGDAAI AALLEGIGDQ VDDDSLMSEK PRIEDLEKSD
 101 DEAPVILIN GLFGQALRLR ASDHIEFFE QSLVNLARD GHLREVLAP
 151 RAISAMLVSR IKVMARLDA EKROPQDRI TLRAGREVD VRVSTLPEIH
 201 GERVYMEVLD KQSLIALDN LGMPAAVLHG LRSCLARPNG IYLTGPTGS
 251 GKTTIVLASL NSLINDSRNI LTVEDPEYA IAGIGGTAIN PRAGITFASG
 301 LRAILRODDP VIMLGEIRDQ ETAQIAVQAS LMGHLVLSL HTNSAVGAVT
 351 RLNDMGIEFP LIASCLRGVL AQRVLRRLCQ CAAVQPLQTA ERDLPELCA
 401 LGSYHAYVC EOCQSGGYVG RLGYEFIEL DAGLIALYD GASEGAMQDY
 451 LADRRRLVA MASDCLARGE TSLAEVLRVV QG

11A:SEQUENCE 1.0
 P1:S52163 - sucrose specific repressor - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
 C:Accession: S52163

R:Bockmann, J.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: S52160
 A:Accession: S52163
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <BOC>
 A:Cross-references: EMBL:X81461; NID:g608705; PID:g608709
 C:Superfamily: lac repressor

S52163 Length: 331 February 11, 2000 15:51 Type: P Check: 8969

1 MASLKDVANL AGVSMYTSR VMHNAESVR ATRDRVLDAI QTLNYPVDSL
 51 ARKMRQGRK PSTLAVLAQD TATTPSVVDI LLAIEQTASE FGNNSEFINI
 101 FSEDNARRA ROLAHRPDG IITTMGLRH ITPESLYGE NYLANCVAD
 151 DPALPSYIPD DYTAQESTQ HLIAGYROP LCFWIPESAL ATGYRQGE
 201 QAVRDAGRDL AEVKQFNAT GDDHYTDLAS LNAHFKSGK PDEVDYICGN
 251 DRAAFVAYOV LIAKGVRIQ DVAVMGFDNL VGVGHFLFPP LTTIQLPHDI
 301 IGRERALHII EGREGRYTR IPCPLIRCS T

11A:SEQUENCE 1.0
 P1:D70044 - transcription regulator GntR family homolog yvoA - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Aug-1999
 C:Accession: D70044
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, J.; Bartsch, L.; Brans, A.; Brann, M.; Bignelli, S.C.; Brox, S.; Brouillet, S.; Bursch, C.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.-J.; Conerton, I.F.; Cummings, R.J.; Daniel, R.A.; Denzot, F.; Devigne, K.M.; Duesterhoef, A.; Ehrlich, S.D.; Emerson, P.T.; Entlian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi, G.; Galleron, N.; Ghim, S.Y.; Glaeser, P.; Goffeau, A.; Gollightly, E.J.; Grandt, G.; Gillespie, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klier, B.; Klier, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogikawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwolik, S.; Prescott, A.M.; Prescetan, E.; Pujic, P.; Purnelle, B.; Rapoport, C.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scioffone, F.; Sekiguchi, J.; Sekowska, A.; Sero, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, R.; Takemaru, K.; Takeuchi, M.; Yamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendool, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wanduit, R.; Wedler, E.; Wedler, H.; Wertzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D70044
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-243 <KUN>
 A:Cross-references: GB:Z99121; GB:AL009126; NID:g3235827; PID:CA815508.1; PID:el186191; PID:g2636016
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yvoA

C:Superfamily: transcription regulator GntR
D70044 Length: 243 February 11, 2000 15:51 Type: P Check: 4428 ..

1 MNINKOSPIP IYYQIMEOLK TOIKNGELQP DMFLPSEREY AEQFISGMT
51 VRQALSLNVLN EGLLYRLKGR GTFVSKPKME QALQGLTSTF EDMRSRGATP
101 GSRIDYOLI DSTEELAIL GCGHPSSTYK ITRVRLANDI PMAISSHIP
151 FELGELNES HFQSSYIDHI ERYNSIPISR AKQELPSSA TTEANILGI
201 OKGAPVLLIK RTTYLNGTA FEHAKSVYRG DRYTFVHYMD RLS

11AA_SEQUENCE 1.0
P1:548826 - NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain pssr precursor - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-Aug-1999
C:Accession: S48826
R:Posson, B.J.; Down, C.G.; Davies, K.M.; Morris, S.C.; Buchanan-Wollaston, V. submitted to the EMBL Data Library, September 1994
A:Description: Nucleotide sequence of a cDNA clone from broccoli (Brassica oleracea L.) with high identity to the pssr subunit of NADH: ubiquinone oxidoreductase
A:Reference number: S48826
A:Accession: S48826
A:Molecule type: mRNA
A:Residues: 1-215 <PDB>
A:Cross-references: EMBL:X82274; NID:9562281; PIDN:CAA57725.1; PID:9562282
C:Genetics:
A:Gene: pssr
A:Genome: nuclear
A:Superfamily: psbc protein
C:Keywords: mitochondrion; NAD: oxidoreductase

548826 Length: 215 February 11, 2000 15:51 Type: P Check: 9647 ..

1 MAMITNTAT RLPLVLOSHR AAASHLATS LPALSPATTP TSYTRBPSS
51 TSAPPPGLSK TAEFVSKVD DLMNWRGRS IMPMTFGIAC CAVENMHTGA
101 ARYDDRFGL IFRSPRQSD CMIVAGLTIN KMAPLRKYV DQMPERAWYI
151 SMSCANGGG YNHSYSYVR GCDRIYVDI YVPGCPFAE ALLYGLLOLO
201 KTRNRKDFL HWMNK

11AA_SEQUENCE 1.0
P1:S75086 - Iron-stress chlorophyll-binding protein - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sl10247
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-Aug-1999
C:Accession: S75086
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naito, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S75086
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1347 <RAN>
A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7948.1; PID:d101881; PID:g1653033
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:
A:Gene: isiA
C:Superfamily: photosystem II chlorophyll a-binding protein psbc
S75086 Length: 342 February 11, 2000 15:51 Type: P Check: 4153 ..

1 MOTYNDTVQ YEMWAGNARF ADQSGLEIA HVAQALTAFA WAGFTLFEI
51 SRFDPTQMG DQGLILPHL AILGWVGDG GQIVDYRPFY VIGSIHLIAS
101 AYLGAALFH TIRAPDLST LKQGRKRFH TWENPQOLGI ILGHILFLG
151 AGALLAKKA MWMGLYDAR TQTVLVSQP TLDPVIYQG QHFASISL
201 EDLVGHTFV GFLIGGIGW HILVPLGMA KYVLSGSA ILSYSLGIA
251 LAGFYAAVFC AVNTLAPPE FYGPLAIKL GIPYFADYV ELPMAHHSR
301 AMLANHEFL AEFPLQGLW HALRALGDFE KRYEQATDSL QT

11AA_SEQUENCE 1.0
P1:A64834 - Probable fimbrial protein-like protein ycbQ precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 26-Aug-1999
C:Accession: A64834
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64834
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-182 <BLAT>
A:Cross-references: GB:AF000196; GB:U00096; NID:g1787169; PIDN:AAC74024.1; PID:g1787170; UMG:p0938
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ycbQ
C:Superfamily: type 1 fimbrial protein
C:Keywords: fimbria
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-182/Product: probable fimbrial protein-like protein ycbQ #status predicted <MAT>

A64834 Length: 182 February 11, 2000 15:51 Type: P Check: 3224 ..

1 MIEKKSVLIT AFITVCATS SYNAADDNAI TDGVTENGK VIAPACTIVA
51 ATKDSVYTLR DVSATKLQTN GOVSGVOIDV PIELKCDIT VTKNAFTFN
101 GTADTGTITA FANQASSDAA TNVALQWYMN DDTTATPDT ETGNILLQDG
151 DQITFEVDY IATGATSGN VNATNHNIN YX

11AA_SEQUENCE 1.0
P1:D71444 - Probable ribosome releasing factor - Chlamydia trachomatis (serotype D strain W3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 26-Aug-1999
C:Accession: D71484
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.F.; Olinger, E.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W. Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.
A:Reference number: A71570; MUID:9900809
A:Accession: D71484
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-179 <RAN>

A:Cross-references: GB:AB001338; GB:AB001273; NID:g3329126; PID:g3329129
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: ftr
C:Superfamily: ribosome releasing factor
D1484 Length: 179 February 11, 2000 15:51 Type: P Check: 1871 ..

1 MTLASERKEM VGVLTFFQKE TRGFRGSKAH PALVEVAYE VYGTMLSD
51 IASISVSDMR QLLSPYAG TVSAISKGL AANLNIQPIV EGAIVRLNV
101 EPREERREV IKOLKRKSE AKVAIRNIR TENDRLKDD NLTEDAVKSL
151 EKKIOELTRK FCKOIEELAK OKAEELATV

!!AA_SEQUENCE 1.0
P1:S74877 - phosphate transport system permease protein C.2, 35K -
Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sir1248
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 26-Aug-1999
C:Accession: S74877
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
Miyajima, N.; Hirose, A.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
T.; Matuno, A.; Muraki, A.; Nakazaki, N.; Nario, K.; Okumura, S.; Shimpou, S.;
Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:title: Sequence analysis of the genome of the unicellular cyanobacterium
Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and
assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S74877
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-328 <RAN>
A:Cross-references: EXBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAAL7838.1;
PID:01018571; PID:g1652920
A:Note: the nucleotide sequence was submitted to the EXBL Data Library, June
1996
C:Genetics:
A:Gene: pstC-2
C:Superfamily: phoW protein
C:Keywords: phosphate transport; transmembrane protein
S74877 Length: 328 February 11, 2000 15:51 Type: P Check: 7310 ..

1 MVEGFRRSE FTLAGGDIT QGNSLTGILD WGFQQLTRLC SLGVVILGW
51 IAWVETDAR PAIAKFGWEF IVSQWDSG QLEGGPIYF GSVVSFIAL
101 IFALPIGLAV AVMTSENLFP APVAVPIAFL VELASIPSEV IIGMGIFVF
151 IPVLAVVOGA LEKYLWLPFI FGTEPEPSPM LVAGVLVTM IIPILASISR
201 DILLSPSPSI RSASMALGAT RWETICVIL PSASSGIITA TIALAGRALG
251 ETMAVTWVG NSNITASL APGYTIPSVL ANQFAVDE LHIGALMTLA
301 LILEVITLGI NSLAVLMVAT IRRGESH

!!AA_SEQUENCE 1.0
F1:A46259 - recombinant protein reca homolog - Arabidopsis thaliana (fragment)
N:Alternate names: recombinase A
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 26-Aug-1999
C:Accession: A46259
R:Centilli, H.; Osman, M.; Grandoni, P.; Jagendorf, A.T.
Proc. Natl. Acad. Sci. U.S.A. 89, 8068-8072, 1992
A:title: A homolog of Escherichia coli RecA protein in plastids of higher
plants.
A:Reference number: A46259; MUID:92390390

A:Accession: A46259
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-438 <CER>
A:Cross-references: GB:M98039; NID:g166840; PIDN:AAA32855.1; PID:g166841
A:Note: sequence extracted from NCBI backbone (NCBIN:112803, NCBI:P:112805)
C:Superfamily: recombinant protein reca
C:Keywords: ATP; DNA binding; DNA recombination; DNA repair; P-loop; SOS
response
F1:41-148/Region: nucleotide-binding motif A (P-loop)
F2:15-220/Region: nucleotide-binding motif B
F3:147/Binding site: ATP (lys) #status predicted
A46259 Length: 438 February 11, 2000 15:51 Type: P Check: 9097 ..

1 DSQVLVSLKL NPSFTPLSPL FPFPCSSFS PSLRSSCYG RLKSPVTIV
51 AAKKLCHKIS SEFDDRINGA LSPDADSRFL DRQKLEAM NDINSGFGK
101 SVTRLGAGG ALVETSSGI LTIDLAGGG LKGRVVEIV GPSSGKTTL
151 ALHAIAEVOK LGNAMLVDA EHAEDPAYSK ALGVVENLI VCPDNGEMA
201 LETADMRCS GAVDLCVDS VSALTFRABI EGEIGMOQG LQARIMSOAL
251 RKSGNASKRA GCTLIFLNOI RYKIGVYGN PEVTSGIAL KFEASVLEI
301 RSAGKIKSK GDEDGLRAR VVQKSKVSR PYQAEFEIM FGEGVKLGC
351 VLDCAIMEV VYKGSWYSY EDORLGQGR KALQHLREMP ALQDEIEKV
401 RLMLDGEVH RSTPLMSSS SSASHREBE EDSLDDFO

!!AA_SEQUENCE 1.0
P1:J00661 - impb protein - Salmonella typhimurium plasmid TP110
C:Species: Salmonella typhimurium
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 26-Aug-1999
C:Accession: J00661
R:Jodavick, D.; Owen, D.; Strike, P.
Nucleic Acids Res. 18, 5045-5050, 1990
A:title: DNA sequence analysis of the imp UV protection and mutation operon of
the plasmid TP110. Identification of a third gene.
A:Reference number: J00659; MUID:90384799
A:Contents: plasmid TP110
A:Accession: J00661
A:Molecule type: DNA
A:Residues: 1-424 <LOD>
A:Cross-references: EXBL:X53528; NID:g47745; PIDN:CAA37608.1; PID:g47748
C:Genetics:
A:Gene: impb
A:genome: plasmid TP110
C:Function:
A:Description: involved in UV protection and mutation
C:Superfamily: umuc protein
C:Keywords: DNA repair; induced mutagenesis; SOS mutagenesis
J00661 Length: 424 February 11, 2000 15:51 Type: P Check: 6434 ..

1 MFALADINSE YASCEKFRP DLREPEYIVL SNNDCCVIAR SPEAKALGR
51 MGQPFQVQO MLEKKIHVF SSNYALYHSM SQRVAVLES LSPAVERYSI
101 DEMFIDLRGI NHCISPEFFG HQLEQVKSX TGLTNGVIA PTKRLASAO
151 WATKQWQFS GVALTAENR NRIKLGLIO PYGEWVGXH RLTEKLALG
201 INTALQAOA NTAIFRKNFS VILERTVREL NSESCISLEE APPAKOOIVC
251 SRSFGRITD KDAHQAVYO YAEFAEKLR GERQCPQYT TFVTSPEAV
301 KERQYNAV EKLPLPTQDS RDIILAAARA LNHVWREGVR YMKGVMLAD
351 FTSSGIAQPG LFDLQPRKN SEKLMTLDE LMQSGKXVW FAGCTAPDW

401 ONKREMLSOC YTTKMDIPL ARLG

!!NA_SEQUENCE 1.0
 PI:F71946 - flagellar motor switch protein - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 C:Title: straiJ99
 C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 26-Aug-1999
 C:Accession: F71946
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deGongne, B.L.; Carnell, G.; Timmimo, P.J.; Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Weiberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.E.; Trust, T.J.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.
 A:Reference number: A71800; MUID:99120557
 A:Accession: F71946
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-343 <ARN>
 A:Cross-references: GB:AE001468; GB:AE001439; NID:94154838; PIDN:AD05900.1; PID:94154847
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: flag
 C:Superfamily: flagellar switch protein flag

F71946 Length: 343 February 11, 2000 15:51 Type: P Check: 9168 ..

1 MATLITRKOR AQDELMSSE KAILILIVG EDTGELIR LDIDSITEIS

51 KOIYVONCTD KOIGAVLEF FFAFOSNOY INTGUEYAR ELTTTLOSE

101 EARKVMDKT KSIOTOKNFA YLGKRPQOL ADLIINHQ TIALIAME

151 APNAETLSY FPDEMAEIS IRMANLEIS PÖYVAKVSY LENKESLTS

201 YKIEVGIGRA VAEIFRQLO KSATTLARI ESDVNLKAGA IKEMFTED

251 IAKIDNFAIR EIKVADKRD LSLAIKISTQ DLTDKFLNNA SRRAAEOYE

301 EMOYLGAVKI KDYVNAORRI IEVOSLOEK GVIOTGEED VIE

!!NA_SEQUENCE 1.0
 PI:A37763 - virD1 protein - Agrobacterium tumefaciens plasmid pTiC58
 C:Species: Agrobacterium tumefaciens
 C>Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 26-Aug-1999
 C:Accession: A37763; S11838; A22666
 R:Wang, K.; Herrera-Estrella, A.; Van Montagu, M.
 J. Bacteriol. 172, 4432-4440, 1990
 A:Title: Overexpression of virD1 and virD2 genes in Agrobacterium tumefaciens enhances T-complex formation and plant transformation.
 A:Reference number: A37763; MUID:90330550
 A:Accession: A37763
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <MAN>
 A:Cross-references: GB:M33673; NID:9142284; PIDN:AAA22110.1; PID:9142285
 R:Rozdowsky, P.M.; Powell, B.S.; Shlitsky, K.; Lin, T.S.; Morel, P.; Zyrjan, E.M.; Steck, T.R.; Kado, C.I.
 Plasmid 23, 85-106, 1990
 A:Title: Molecular characterization of the vir regulon of Agrobacterium tumefaciens: complete nucleotide sequence and gene organization of the 28.93-kbp regulon cloned as a single unit.
 A:Reference number: S11825; MUID:90301800
 A:Accession: S11838
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-147 <ROC>
 A:Cross-references: EMBL:J03320; NID:9154781; PIDN:AAA91603.1; PID:9154795
 R:Hagiya, M.; Close, T.J.; Tait, R.C.; Kado, C.I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 2669-2673, 1985

A:Title: Identification of pTiC58 plasmid-encoded proteins for virulence in Agrobacterium tumefaciens.
 A:Reference number: A94037; MUID:85190558
 A:Accession: A22666
 A:Molecule type: DNA
 A:Residues: 1-57, 'P', '59-105', 'SPAS', 'HAG'
 A>Note: the authors translated the codon ATA for residue as Met, ATA for residue 79 as Met and ATA for residue 83 as Met
 C:Genetics:
 A:Gene: virD1
 A:Genome: plasmid
 C:Superfamily: virD1 protein

A37763 Length: 147 February 11, 2000 15:51 Type: P Check: 2312 ..

1 MGOGRPSS DIANORECV KYEGKRVST RLRSAREYF SHQARLIGLS

51 DSMAIRAVR RIGGLEIDA ETRHREAIL QSIGTLSSNI AALLSYAEN

101 PTMDLEALRA ERIAFKSPA DLDGLRSTL SVSRRIDGC SLIKDAL

!!NA_SEQUENCE 1.0
 PI:A25063 - hypothetical virD1 protein - Agrobacterium tumefaciens plasmids
 C:Species: Agrobacterium tumefaciens
 C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 26-Aug-1999
 C:Accession: A29826; A25063
 R:Jayaswal, R.K.; Veluthambi, K.; Gelvin, S.B.; Silghom, J.L.
 J. Bacteriol. 169, 5035-5045, 1987
 A:Title: Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molecules in Escherichia coli by a virD-encoded border-specific endonuclease from Agrobacterium tumefaciens.
 A:Reference number: A91846; MUID:88032822
 A:Accession: A29826
 A:Molecule type: DNA
 A:Residues: 1-147 <JAY>
 A:Cross-references: EMBL:M17989; NID:9142288; PIDN:AAA22113.1; PID:9142289
 A>Note: plasmid pTiA6
 R:Yanofsky, M.F.; Porter, S.G.; Young, C.; Albright, L.M.; Gordon, M.P.; Nester, E.W.
 Cell 47, 471-477, 1986
 A:Title: The virD operon of Agrobacterium tumefaciens encodes a site-specific endonuclease.
 A:Reference number: A90886; MUID:87028239
 A:Accession: A25063
 A:Molecule type: DNA
 A:Residues: 1-147 <IAN>
 A>Note: plasmid pTiA6NC
 C:Genetics:
 A:Gene: virD1
 A:Genome: plasmid
 C:Superfamily: virD1 protein
 C:Keywords: endonuclease

A25063 Length: 147 February 11, 2000 15:51 Type: P Check: 4145 ..

1 MSKHRTSS ETAINORSL NVEGFRVSA RLRSAREYF SYQARLIGLS

51 DSMAIRAVR RIGGLEIDA HTRKMEAIL QSIGTLSSNV SMDLSAYED

101 PRSDLAVRD ERIAFGEAFA ALDGLRSTL SVSRRIDGC SLIKGAL

!!NA_SEQUENCE 1.0
 PI:S06883 - virD1 protein - Agrobacterium rhizogenes plasmid pTiA4b
 C:Species: Agrobacterium rhizogenes
 C>Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 26-Aug-1999
 C:Accession: S06883
 R:Hirayama, T.; Muranaka, T.; Okawa, H.; Oka, A.
 Mol. Gen. Genet. 213, 229-237, 1988
 A:Title: Organization and characterization of the virD genes from Agrobacterium rhizogenes.
 A:Reference number: S06883; MUID:89039712
 A:Accession: S06883
 A:Molecule type: DNA

A:Residues: 1-147 <HIR>
A:Cross-references: EMBL:X12867; NID:g38995; PIDN:CAA31350.1; PID:g38999
C:Genetics:
A:Gene: vird1
A:Genome: plasmid
C:Superfamily: vird1 protein

S0683 Length: 147 February 11, 2000 15:51 Type: P Check: 2761 ..

1 MSGSKRTSS DAINRGVGA TVEGRVYST RLRSAYESP SHQARLGLS
51 DSAIRAVR RIGGFLEIDA ETRKMEIL LSTGTLSSNI AALLSAYEN
101 PTMDLELRA ERIRAGESTF DLDELRSIL SVSRRRIDGC SMLKDSL

!!A_SEQUENCE 1.0

P1:S36564 - E2 protein - human papillomavirus type 45

C:Species: human papillomavirus type 45

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S36564

R:Delius, H.; Hofmann, B.

Submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36569

A:Accession: S36564

A:Molecule type: DNA

A:Residues: 1-368

A:Cross-references: EMBL:X74479; NID:g397022; PIDN:CAA52576.1; PID:g397026

C:Superfamily: papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

S36564 Length: 368 February 11, 2000 15:51 Type: P Check: 5828 ..

1 MKMOTPRESL SERLSLQDK ILDRYENDSK DINSQISTYQ LIRLENALIF
51 TAREHGITKL NHOVPPINI SKSKRAKAE LQALKGLAQ SKYNNENETL
101 QDTCEELMNT EPSQCFKGG KTVHYFDGN KDNCKMYVW DSIYITETG
151 IMDKTAQVCS YWGVYYIKDG DTYVVOFKS ECEKYGNSNT MEVOYQGANVI
201 DCNDSMCSSTS DDTVSATQIV RQLOHASTST PRTASVGTFR PHIOPTATKR
251 PROCGLEQHG HGRVNTVHN PLLCSTSNM KRRKVCSTN PTIHLKGDK
301 NSLKCLNRYRL RKYADHYSEI SSTWHTGNCN KNTGILLTVY NSEVQRNFTL
351 DVTIRPSVQ ISVGKMTI

!!A_SEQUENCE 1.0

P1:S36576 - E2 protein - human papillomavirus type 52

C:Species: human papillomavirus type 52

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S36576

R:Delius, H.; Hofmann, B.

Submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36576

A:Molecule type: DNA

A:Residues: 1-368

A:Cross-references: EMBL:X74481; NID:g397038; PIDN:CAA52588.1; PID:g397042

C:Superfamily: papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

S36576 Length: 368 February 11, 2000 15:51 Type: P Check: 4491 ..

1 MESIPARLNA VOBKILDLVE ADSNDLNAQI EHKILIRMEC VLFYAKBELG
51 ITHIGHQVVP PMAVSKAKAC QALIELQALE ALKKTQYSTD GWTLOQTSLE
101 MMRAEPQKTF KKHGTTIVQ YDNDKNNMTD YTNMKETIYL GECECTIVEG

151 QVDYIGLYW CDGEKIYFVK FSNDAKOYV TGWMEVHNG QVIVCAPSVS
201 SNEVSTTERA VHLCTETSKT SAVSVGAKDT HLOPPQRRR PDVDSRNTK
251 YPNNLRGQO SYDSTTRGLV TATECTNKR VAHTCTAPI IHLKDPNSL
301 KCLRIRYKTH KSLYQOIST WHMTSNECTN NKLGIYTVY SDETOROOFI
351 KTVKIPRTVQ VIGVMSL

!!A_SEQUENCE 1.0

P1:S36552 - E2 protein - human papillomavirus type 3

C:Species: human papillomavirus type 3

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S36552

R:Delius, H.; Hofmann, B.

Submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36552

A:Molecule type: DNA

A:Residues: 1-383

A:Cross-references: EMBL:X74462; NID:g397005; PIDN:CAA52472.1; PID:g397009

C:Superfamily: papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

S36552 Length: 383 February 11, 2000 15:51 Type: P Check: 4819 ..

1 METLANRLDV CODKILELVE KSDKLEBDQI MMQOLRLQO ALVYKARECG
51 LTHIGHQVVP PLSTYRKAR SAIEVHVSQ QLOHSHADQ PWTLRDTSRE
101 MMDTVPRKCM KRGGLTVFR YDGENKAMC YQWREITVQ NYTDNNWYK
151 AGLVSHGLY YMHGQKTFY VKFKDARVY GDTGWDVAV GGKVIHHSF
201 DPVSTRELP APGLYACTT QAPLOAVGA SEQPEKORR LETVYGEQOQ
251 QOQOQOQOQOQ HTQTPAQTT ERARQPLDQD RTDRDITCP HPIGHRSDDP
301 CVPYHLRGD PNCLEFRYR LNKGNKLYS RSTSTWRMS C ESENOCAVYT
351 IWTYSYGORE AFLSTYKVP GIVYLGHMS MET

!!A_SEQUENCE 1.0

P1:S36535 - E2 protein - human papillomavirus type 10

C:Species: human papillomavirus type 10

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S36535

R:Delius, H.; Hofmann, B.

Submitted to the EMBL Data Library, August 1993

A:Description: Primer-directed sequencing of human papillomavirus types.

A:Reference number: S36469

A:Accession: S36535

A:Molecule type: DNA

A:Residues: 1-376

A:Cross-references: EMBL:X74465; NID:g396901; PIDN:CAA52492.1; PID:g396905

C:Superfamily: papillomavirus E2 protein

C:Keywords: DNA binding; early protein; transcription regulation

S36535 Length: 376 February 11, 2000 15:51 Type: P Check: 273 ..

1 METLANRLDA CODKMLELVE KSDKLEBDQI TMHHLRIVEN ALLYKARECG
51 LTHIGHQVVP PLSTYRKAR NAIEVHALQ QLOESAYHE PWTLRDTSRE
101 MMDTAPKCM KRGGLTVFR YDGENKAMC YQWREITVQ NYSDDRWYK
151 PGKYSIEGLY YTHENNTIY VNFKODACYV GETGKREIVY GGKVIHHDAT
201 DPVSTREIS TPQVCTSTNT TPASTQAVG ASEGEQKQO RUEAVDGOHQ
251 QORSGKSDST OKAERBAGQ VSDRRLDQD TSSAHEVRHP SDPCAPYIH

```

301 LAGDENSILKC FRYRLHNGKR KLYSRSSSTW RMCSESENOA AFVTLMTSD
351 TORTEFLNV KVPPGIVIL GVMSEF

!!A:SEQUENCE 1.0
P1:A44215 - EUS1 protein - equine herpesvirus 1 (strain Kentucky A)
C:Species: equine herpesvirus 1
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C:Accession: A44215
R:Breedem, C.A.; Yalamanchili, R.R.; Colle III, C.F.; O'Callaghan, D.J.
Virology 191, 649-660, 1992
A:Title: Identification and transcriptional mapping of genes encoded at the
IR/US junction of equine herpesvirus type 1.
A:Reference number: A44215; MUID:93079867
A:Accession: A44215
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <BRE>
A:Cross-references: GB:M80429
C:Superfamily: herpesvirus US2 protein

A44215 Length: 303 February 11, 2000 15:51 Type: P Check: 5996

1 MGVLLIVVT VVDRHKLAPN SSIDVDGHLW EFLSRQCFVL ASEPLGIPIY
51 VRSADYRFS SLLTLPRKC RPIYTRGDT AIALDRNGV YHEDRMGVSI
101 EMLSVLSGN HLNSSLINQ PYHLVVGAA DICKPVFDLI PGPFRVYAE
151 IADFHSMQ PPFVCGKLE TTPMTVEHN HPLKLAAG EDIVVGECHF
201 SKHSNSLVH PPTVNVVIA VVDPARLEI PAFGRPLPRP RPSGGRAP
251 RRRRARAPA RSTAAATRP RCPDRAPAA RAGDVTWME RLVGVGTGT
301 STR

!!A:SEQUENCE 1.0
P1:F43674 - US7 protein - human herpesvirus 2 (strain HG52)
C:Species: human herpesvirus 2
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: F43674
R:McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.
J. Gen. Virol. 68, 19-38, 1987
A:Title: DNA sequence and genetic content of the HindIII 1 region in the short
unique component of the herpes simplex virus type 2 genome: identification of
the gene encoding glycoprotein g and evolutionary comparisons.
A:Reference number: A43674; MUID:87111457
A:Accession: F43674
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <MCG>
A:Cross-references: EMBL:X04798; NID:959900; PIDD:CAA28485.1; PID:959906
C:Superfamily: herpesvirus US7 protein

F43674 Length: 372 February 11, 2000 15:51 Type: P Check: 501

1 MPERSLQGLA ILGLWCATG LVRGPTVSL VSDSLVDAGA VDGQGVFED
51 LRVFGLHEV GAQVPHNTY DGILEFHP LGNHCPRYVH VVLTACPRR
101 PAVAFILCRS THHASPAPV TLEGLARCP LLRVATARD YAGLYLVRYW
151 VGSATNSRF VLGVALSANG TFWYNGSDYG SCDPAQLPFS APRLGSSVY
201 TPGRSRTPP RTTTPSSPR DTPAPGDTG TPAPASGEIA PPNSTRASAE
251 SRHLVTAQV IQAIPASII AFVLSGSCIC FIHQORRYR RPRGQIYNPG
301 GVSACVNDAA MARLGAELRS HPNTPPKPR RSSSSTWPS LTSIAESEP
351 GPVVLVSVP RPSGPTAQ EV

!!A:SEQUENCE 1.0
P1:S58346 - coat protein V1 - tomato yellow leaf curl virus
C:Species: tomato yellow leaf curl virus
C>Date: 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 20-Sep-1999
C:Accession: S58346
R:Hong, Y.; Harrison, B.D.
submitted to the EMBL Data Library, February 1995
A:Description: Nucleotide sequences from tomato leaf curl viruses from
different countries: evidence for three geographically separate branches in
evolution of the coat protein of whitefly-transmitted geminiviruses.
A:Reference number: S58346
A:Accession: S58346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <HON>
A:Cross-references: EMBL:Z48182; NID:9944838; PIDD:CAA88227.1; PID:9944839
C:Superfamily: cassava latent virus coat protein

S58346 Length: 256 February 11, 2000 15:51 Type: P Check: 2806

1 MSRRPADMTI GQPVSKYRRL LSSISPYSKR AAVRTVGTG KEMANRPVN
51 RKPWFYMER GPDYRGCEG PCKVQSFESR HDIHIKVM CISDVTGTC
101 LTRVGRKFC VASVYVLGKI WMDENIKTN HINSVFFLV RDRRPVDRQ
151 DFGDVEMED NEPTATYKN MHRDROYLR KWHATVTGGQ YASKEQALVK
201 EFVKVNNYV YNQEAGRYE NHSENALMTY MACTHASNPV YATLKIRIYF
251 YDSVTN

!!A:SEQUENCE 1.0
P1:J01248 - hypothetical 12K protein - chrysanthemum virus B
C:Species: chrysanthemum virus B
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Feb-1997
C:Accession: J01248
R:Levay, K.; Zaytsev, S.
J. Gen. Virol. 72, 2333-2337, 1991
A:Title: Nucleotide sequence and gene organization of the 3'-terminal region of
chrysanthemum virus B genomic RNA.
A:Reference number: J01248; MUID:92013948
A:Accession: J01248
A:Molecule type: genomic RNA
A:Residues: 1-106 <LEV>
A:Note: the authors translated the codon ATC for residue 57 as Tyr
C:Superfamily: barley stripe mosaic virus 14K protein
C:Keywords: transmembrane protein

J01248 Length: 106 February 11, 2000 15:51 Type: P Check: 566

1 MFLTPPDHT KVLVAAGL SIVASILYS RNTLPVGGH SHLPHRGVY
51 KDGTKIYVG GPRKLSLEG GENLVPQWF LVLLSNALF LLSRSGHRR
101 VCGSCH

!!A:SEQUENCE 1.0
P1:S41285 - coat protein - sweet potato G virus (fragment)
C:Species: sweet potato G virus
C>Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 20-Sep-1999
C:Accession: S41285
R:Colinet, D.; Kummerl, J.; Lepoint, P.
submitted to the EMBL Data Library, December 1993
A:Reference number: S41285
A:Accession: S41285
A:Molecule type: DNA
A:Residues: 1-335 <COL>
A:Cross-references: EMBL:X76944; NID:9439687; PID:9439688
C:Superfamily: tobacco etch virus genome polypeptide; DEAD/H box helicase
homology
C:Keywords: coat protein

```


541285 Length: 355 February 11, 2000 15:51 Type: P Check: 3148

1 SAEIYDACK TGNNGRGR GTVPPPPPP GAPRTGDI PP AVOTGPPPG
51 ASKPPPIIE ILQPSPTK ALREARKAP AIPDSRGVD ISQIPSTPG
101 RQOTMPTPO RTSTEVDRD VNAGTVGTI VPRQITHSK KTAPMANGRI
151 VVNDHLITV DPEQTSLSNT RATOEFNFM YEGVEDYGV NDBOMGILLN
201 GLAWCIENG TSPNINQMV MMDGBOVY PIKPLDNAV PTERQIMTHF
251 SDIAEYTER RNRIKAYMP YGLQRNLTDM SIARYAFDF ELHSNTPVRA
301 REAHQMKAA ALKNAQRNF GLDGNVSTQE EDTERHTTD VTNINHLG
351 MRGVQ

!!AA_SEQUENCE 1.0
PI:E71669 - preprotein translocase secy chain (secy) RP639 - Rickettsia
prokazekii
C:Species: Rickettsia provazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 26-Aug-1999
R:Accession: E71669
R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.;
Almark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.;
Kurland, C.G.
A:Title: The genome sequence of Rickettsia provazekii and the origin of
mitochondria
A:Reference number: A71630; MID:99039499
A:Accession: E71669
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-433 <N>
A:Cross-references: GB:AJ235272; GB:AJ235269; MID:93861033; PID:e1342923;
PID:93861179
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: secy, RP639
C:Superfamily: preprotein translocase secy

E71669 Length: 433 February 11, 2000 15:51 Type: P Check: 3620

1 MGNFSKSS NDVNRILFT LFMILICRFG SPIPIGIDS IALNSVAENK
51 QFQILGMNM LSGSLGRMS IFALAIWPI TASIIQLMS VAKPLENLK
101 KEGTGKRI NQLSRYLTV LASFOAYGVA LLSMVTNT GPVILLAGEF
151 FRVTVITLV VGTILMWIG EQITRGIGN GTSIIIFGI ISGVPSAIIIS
201 MEISRKAL SPLAIYVCI GVVLAIIII FFEKQRRL VOYPRKQVGN
251 KIYGEATHM PLKNTSGVI PRIFASSILL FPTILASNS SNTDYSMLT
301 YLHGKRPVY ILVYVLMF ESFFYAIYF NSEETANLR KIGAYIPGKR
351 PKNNSDFEV YILRLVIGI GYLSVTCYI PELIMNKYVI SLSLGTSLF
401 IVVNVLDTM TQTYTFSS KTEGLMKRIK LKN

!!AA_SEQUENCE 1.0
PI:G69632 - transcription repressor glcr - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
R:Accession: G69632
R:Kunst, R.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,
V.; Betteiro, M.G.; Bessieres, P.; Bolocin, A.; Borcherdt, S.; Boriss, R.;
Bourstier, L.; Brans, A.; Brann, M.; Brignelli, S.C.; Bron, S.; Bruggliet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,
J.O.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denicot, F.; Devine, K.M.;

Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Feriati, E.
A:Title: The complete genome sequence of the gram-positive bacterium Bacillus
subtilis.
A:Reference number: A69580; MID:96044033
A:Accession: G69632
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-258 <N>
A:Cross-references: GB:Z99122; GB:AL009126; MID:92636029; PID:e184536;
PID:92636155
A:Experimental source: strain 168
C:Genetics:
A:Gene: glcr
C:Superfamily: regulatory protein glcr

G69632 Length: 258 February 11, 2000 15:51 Type: P Check: 4284

1 MYOERLVAI LDFLKHNRIT TEOITCTILO VSRDTRARDL VKLEQNALIT
51 RTRGAILPT VHOKIOSYSG RKTIVSEKN KIGSLASLI HOGDRIYIDA
101 STTVQACAKH LNAVDCTVIT NSINLADVLIS DKEIEIYLL GGRLEKHF
151 IYSSVIEKL SSYHVDKALI GVGISEHGI TINHEEDGMV KKKMIQAKO
201 VIALADHSLK GSTSFQYAE LNEIDLITD RLPRQAFCD LDRNGVELLV
251 TEODEGRD

!!AA_SEQUENCE 1.0
PI:F69260 - nitrate ABC transporter, permease protein (ntrB-1) homoIog -
Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
R:Accession: F69260
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum,
K.A.; Dodson, R.J.; Gaitan, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.;
Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush,
J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirtness, E.F.; Dougherty, B.A.;
McKenney, N.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.R.;
Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.;
McDonald, L.
A:Title: The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
A:Reference number: A69250; MID:98049343

A:Accession: F69260
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-244 <KUN>
A:Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PID:g2650562;
TIGR:AF0086
C:Superfamily: Synechococcus nitrate transport protein ntb
F69260 Length: 244 February 11, 2000 15:51 Type: P Check: 6950 ..

1 M0ASKADYKG YLTITSLT IMAAASYLO NPALNPDPV LALLTKEL
51 ITHAVSLIR VVYSLALSL VALPAGILSR ERYVDATISF IYILYIPR
101 IVLPFLYLL FGIGDLSRVL LIATLEFQI AVTRDPAKO VSDYVYSIL
151 SLGASKIDY RHYIIPAVMP KILPALRISI GNAIVAFRA ESFAITSGLG
201 YLIDMSERA DYTMYAAT TMAILGELY VIVESAERV CRWL

!!AA-SEQUENCE 1.0
PI:A70001 - ABC transporter (ATP-binding protein) homolog yfsc - Bacillus
subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-May-1999
C:Accession: A70001
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,
V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.;
Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, K.M.;
J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devane, K.M.;
Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entlian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi,
A.; Galleron, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Goldlighty, E.J.; Grandi,
G.; Galspehl, G.; Guy, B.J.; Haga, K.; Haeisch, J.; Harwood, C.R.; Henaut, A.;
Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.;
Joris, B.; Karmata, D.; Kasahara, Y.; Kletter-Blanchard, M.; Klein, C.;
Kobayashi, Y.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;
Lapilus, A.; Lardinois, S.
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,
B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,
S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato,
T.; Scanlon, E.
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,
A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.;
Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka,
T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.;
Vannier, F.; Vassaret, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.;
Wetzneger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A70001
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-253 <KUN>
A:Cross-references: GB:299119; GB:AL009126; NID:g2635411; PID:el185911;
PID:g2635522
A:Experimental source: strain 168
C:Genetics:
A:Gene: yfsc
C:Superfamily: ATP-binding cassette homology
C:Keywords: ATP; P-loop
F:23-219/Domain: ATP-binding cassette homology <ABC>
F:40-47/Region: nucleotide-binding motif A (P-loop)

A70001 Length: 253 February 11, 2000 15:51 Type: P Check: 6987 ..

1 MVLNENKIR KSYCNKLNKO EYLGKIDHI EKGFVYSIMQ ASGSGKTLT
51 NVISSDQYS HGTHINGD MTKAMKOLA EFRKHGFI FQDYMILDTL
101 TVKENILPL SITLKSKEA NKRFEVAKI LGEIENDKY PNEISGOKO
151 RISAGRAFIH DSIIFADEP TGAIDSKAS DLNKLSQLN QKRNATIMV
201 THDVAASYC GRVIFIKDQ MTQLNKGO DROFFQDIM KQGVILGGVO
251 HEH

!!AA-SEQUENCE 1.0
PI:A70039 - ABC transporter (ATP-binding protein) homolog yfvr - Bacillus
subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-May-1999
C:Accession: A70039
R:Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,
V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.;
Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, K.M.;
J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devane, K.M.;
Duesterhoft, A.; Ehrlich, S.D.; Emerson, P.T.; Entlian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi,
A.; Galleron, N.; Gilm, S.Y.; Glaser, P.; Goffeau, A.; Goldlighty, E.J.; Grandi,
G.; Galspehl, G.; Guy, B.J.; Haga, K.; Haeisch, J.; Harwood, C.R.; Henaut, A.;
Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Itaya, M.; Jones, L.;
Joris, B.; Karmata, D.; Kasahara, Y.; Kletter-Blanchard, M.; Klein, C.;
Kobayashi, Y.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;
Lapilus, A.; Lardinois, S.
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,
B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,
S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato,
T.; Scanlon, E.
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,
A.; Seror, S.J.; Serro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.;
Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka,
T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.;
Vannier, F.; Vassaret, A.; Viari, A.; Wandut, R.; Wedler, E.; Wedler, H.;
Wetzneger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A70039
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-301 <KUN>
A:Cross-references: GB:299121; GB:AL009126; NID:g2635827; PID:el186097;
PID:g2635922
A:Experimental source: strain 168
C:Genetics:
A:Gene: yfvr
C:Superfamily: ATP-binding cassette homology
C:Keywords: ATP; P-loop
F:21-206/Domain: ATP-binding cassette homology <ABC>
F:38-45/Region: nucleotide-binding motif A (P-loop)

A70039 Length: 301 February 11, 2000 15:51 Type: P Check: 1714 ..

1 MKKEAVTVSN VTKHFORKTA VNNISFEIER GELAILBPN GAGRTVSM
51 ILGLIKBSG EIKLENRPD DQVRKIGV MGEVSVMPG LKVDLIELF

101 RSYPNPLSM KELVSLTALT KEDLTKRAEK LSGGOKRRLS FALALANPE
 151 LLIDEPVVG MDTSRRHRM QTINGLSOG KTIIFSRHYL QADDAQORI
 201 LFTFEGOLVA DSGPMOIRSR IQKQSVSTLL HSESELERLS CPEYERVIIH
 251 EHERTTIQTS NTDKVLALIF QENIHARDIR IEQATIDEAF ROLADGNREA
 301 M

!!AA_SEQUENCE 1.0
 P1:G69260 - nitrate ABC transporter, ATP-binding protein (ntrC-1) homolog -
 Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 28-May-1999
 C:Accession: G69260
 R:Kienk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kellavagge, A.R.; Graham, D.E.; Kyriides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kinkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Bagger, J.H.; Glodex, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Ullrich, T.; Cotton, M.D.; Springs, T.; Attleach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
 A:Reference number: A69250; MUID:98049343
 A:Accession: G69260
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-243 <KDE>
 A:Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PID:g2650561; TIGR:AF0087
 C:Superfamily: ATP-binding cassette homology
 C:Keywords: ATP; P-loop
 F:17-205/Domain: ATP-binding cassette homology <ABC>
 F:34-41/Region: nucleotide-binding motif A (P-loop)
 G69260 Length: 243 February 11, 2000 15:51 Type: P Check: 9269 ..

1 MITVKGISKF FGLKALDGV SFVEVDGETC AITGPGCGK STLLIMAGL
 51 LKPSGEVLV DRAVNSPLK NAALLIDDFG LFPKTYVDN VALGKIRGF
 101 SRAERERTV ALLEKFLGK FEKSYKOLS GGMKORVAA RAIAEPOLL
 151 LMDPLSLD ALSREMONF LNLKKEKTK TMLVTHSIE EAVFLGRKIV
 201 VLTERPGRVK AAVDNRDAG ESYREVEVF ERCKLRQTI RAT

!!AA_SEQUENCE 1.0
 P1:S56335 - hisP-like nucleotide binding protein phnK - Escherichia coli
 N:Alternate names: phosphonates transport ATP-binding protein phnK
 C:Species: Escherichia coli
 C:Date: 26-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 28-May-1999
 C:Accession: S56335; H65216; C35719
 R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.
 A:Reference number: S56314; MUID:95334362
 A:Accession: S56335
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <BUR>
 A:Cross-references: EMBL:U14003; NID:g1263172; PID:g3536941
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Colado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: H65216
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <BLAT>
 A:Cross-references: GB:AE000482; GB:U00096; NID:g2367349; PID:g1790535; UWGP:D4097
 A:Experimental source: strain K-12, substrain MG1655
 R:Chen, C.M.; Ye, Q.Z.; Zhu, Z.; Wanner, B.L.; Walsh, C.T.
 J. Biol. Chem. 265, 4461-4471, 1990
 A:Title: Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psid) genes involved in alkylphosphate uptake and C-P lyase activity in Escherichia coli B.
 A:Reference number: A35718; MUID:90170953
 A:Accession: C35719
 A:Molecule type: DNA
 A:Residues: 1-46 / 48-252 <CHE>
 A:Cross-references: GB:J05260; NID:g147192; PID:g147206
 C:Genetics:
 A:Gene: phnK
 C:Superfamily: ATP-binding cassette homology
 C:Keywords: ATP; P-loop
 F:21-222/Domain: ATP-binding cassette homology <ABC>
 F:38-45/Region: nucleotide-binding motif A (P-loop)
 S56325 Length: 252 February 11, 2000 15:51 Type: P Check: 1039 ..

1 MNPLSLVNN LTHLYAPGKG FSDVSEDLMP GEVLGIVGES GSGKTTOLKS
 51 ISARLTPQOG EHYENRSYLY AMSEADRRL LRTEGVVHO HPLDGLRROV
 101 SAGNIGERL MATGARHNGD IRATAQKWE EVEIPANRID DLPTFGSGM
 151 QQRQIARNL VTHPKLVEMD EPTGGLDVSV QARLDDLKG LVLELNIAV
 201 IVHDLGVAR LLAIDLIVMK QGVVSEGLT DAVLDPPHNP YTOLVSSVL
 251 ON

!!AA_SEQUENCE 1.0
 P1:S73545 - ATP-dependent RNA helicase dead - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
 N:Alternate names: hypothetical protein C12.ori450
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 20-Sep-1999
 C:Accession: S73545
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
 A:Reference number: S73327; MUID:97105885
 A:Accession: S73545
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-450 <HIM>
 A:Cross-references: EMBL:AE000022; GB:U00089; NID:g1673882; PID:NAB95867.1; PID:g1673885
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Gene: dead
 A:Genetic code: SGC3
 C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
 C:Keywords: ATP; P-loop
 F:47-342/Domain: DEAD/H box helicase homology <DEAD>
 F:47-54/Region: nucleotide-binding motif A (P-loop)
 F:150-155/Region: nucleotide-binding motif B
 F:154-157/Region: DEAD motif

```

573545 length: 450 February 11, 2000 15:51 Type: P Check: 2944
1  MSDTFNELGV SPALITATLKD NNINOPTTIO QLAIPQLOH QNLIVHSPTG
51  TGTAAVEGIP VIETLLKKPS KGTOTLVVA PYRELAQIK TTEINFKAHT
101  HLMVSLIGG IPIWOQLKOL ENOEIYVGT MGRVMDLLER GVIFEELEH
151  LIIDVDVLM DRGFRRKFLD LLSRIEFEG IAVYSASVNE ETIEFAKOIT
201  KNGIFLAPE LKONAPEDPN KLIDQFVCL FSNRRKQALY SLVSTQFAS
251  IYPCDTRKL VDELCIFLR NDVXTYPLHG DKAQFIREN LKLANPTAP
301  IVLTVDLIG RGIHVEGDM VVNSACVNF EYIHRMGT GRNNHKGSCI
351  TFCISHKOA FLKLEQVND KRISPLRPM LRILPLKCT OPRKGRSLD
401  SVOKIYNPR SNGTRRVPPL AGDLFSRMR OPRDMQKNK LHSDMQSNM

11AA_SEQUENCE 1.0
P1:566920 - Probable RNA helicase CAS/6 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein O2774; protein YOR046c
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 20-Sep-1999
C:Accession: S66920; S66929; D34848
R:Landt, O.; Hiesel, R.; Unsel, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66907
A:Accession: S66920
A:Molecule type: DNA
A:Residues: 1-482 <LAN>
A:Cross-references: EMBL:Z74954; NID:G1420174; PIDN:CAA9237.1; PID:E252333;
PID:G1420175; MIPS:YOR046c
R:Bohn, C.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Dang, D.V.; Valens, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66929
A:Accession: S66929
A:Molecule type: DNA
A:Residues: 1-246 <BOH>
A:Cross-references: EMBL:Z74954; MIPS:YOR046c
R:Chang, T.H.; Arenas, J.; Abelson, J.
Proc. Natl. Acad. Sci. U.S.A. 87:1571-1575, 1990
A:Title: Identification of five putative yeast RNA helicase genes.
A:Reference number: A34848; M01D:90180368
A:Accession: D34848
A:Molecule type: DNA
A:Residues: 237-425 <CHA>
C:Genetics:
A:Gene: SGD:DBP5
A:Cross-references: SGD:S0005572; MIPS:YOR046c
A:Map position: 15R
C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology
C:Keywords: ATP; P-loop
F:138-429/Domain: DEAD/H box helicase homology <DEAD>
F:138-145/Region: nucleotide-binding motif A (P-loop)
F:235-240/Region: nucleotide-binding motif B
F:239-242/Region: DEAD motif

566920 length: 482 February 11, 2000 15:51 Type: P Check: 2958
1  MSDTRDPAD LASLKIDNE KEDTSEVSTK ETVASQPEKT ADSIRPAEKL
51  VKVEEKKTK QEDSNLTISSE YEAVKVLADI QADNSPLYS AKSFDELGA
101  PELLGITYAM KFOKPSKIOE RALPPLLNHP PRNNIAOSQS GKGKAAASL
151  TMLTRVNEPD ASPOAICLAP SRELARQOLE VVQDMGKFK ITSQILVDS
201  FEKNQINAO VIVGTPTVL DLMRRKMLQI QKTIIFVDE AAMMDQOGL

251  GDQCIRKRF LPRDQVLVE SATPADAVRQ YAKKIIVPAN TLEQTEVNV
301  VDAIKQLYMD CKNADKEDV LTELGLMTI GSGIIFVATK KTANVLXGKL
351  KSEGEVSIIL HGDLOTQERD RLIDDFREGR SKVLTITNVL ARGIDITFVS
401  MVVNYDPLET ANGQADPATY IHRIGRTGRF GRGVAISFV HDKNSFNLS
451  AIQRYEGDIE MREVPDMD EVERIVKVL KD

11AA_SEQUENCE 1.0
P1:A38900 - T-cell-specific transcription factor-1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 08-Sep-1997
C:Accession: A38900
R:Oosterweel, M.; van de Wetering, M.; Dooljes, D.; Klomp, L.; Winoto, A.;
Georgopoulos, K.; Meljink, F.; Clevers, H.
J. Exp. Med. 173, 1133-1142, 1991
A:Title: Cloning of murine TCF-1, a T cell-specific transcription factor
interacting with functional motifs in the CD3-epsilon and T cell receptor alpha
enhancers.
A:Reference number: S17398; M01D:91217625
A:Accession: A38900
A:Molecule type: mRNA
A:Residues: 1-272 <OOS>
A:Experimental source: T-cell, cell line
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: transcription factor
F:133-23/Domain: DNA binding #status predicted <DNA>
F:154-229/Domain: HMG box homology <HMG1>

A38900 length: 272 February 11, 2000 15:51 Type: P Check: 3880
1  MYKETVYSAF NLMPYPAS GAGQHPQOP PLNKRQOPP HGVPLSLY
51  EHFSSPHPT APADISQKOG VHRPLQTPDL SGFYSLTGS MGQLPHTVSM
101  FTHPSLMGS GVPGHPAIP HPAIVPSSGQ QELQPYDRNL KTOAPEKAK
151  EAKRPVIKRP LNAFMLYMKR MKRATIACT LKESAINOI LGRVHALSR
201  EQQAYEELA RKEROLHML YPGWSARDNY GKRRRSREK HOESTTGKR
251  NAFGTPEKA AAPAFLEPMT VL

11AA_SEQUENCE 1.0
P1:JH0402 - T-cell-specific transcription factor-1D - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 08-Sep-1997
C:Accession: JH0402
R:Oosterweel, M.; van de Wetering, M.; Dooljes, D.; Klomp, L.; Winoto, A.;
Georgopoulos, K.; Meljink, F.; Clevers, H.
J. Exp. Med. 173, 1133-1142, 1991
A:Title: Cloning of murine TCF-1, a T cell-specific transcription factor
interacting with functional motifs in the CD3-epsilon and T cell receptor alpha
enhancers.
A:Reference number: S17398; M01D:91217625
A:Accession: JH0402
A:Molecule type: mRNA
A:Residues: 1-279 <OOS>
A:Experimental source: T-cell, EL-4
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: transcription factor
F:154-229/Domain: HMG box homology <HMG1>

JH0402 length: 279 February 11, 2000 15:51 Type: P Check: 9457
1  MYKETVYSAF NLMPYPAS GAGQHPQOP PLNKRQOPP HGVPLSLY
51  EHFSSPHPT APADISQKOG VHRPLQTPDL SGFYSLTGS MGQLPHTVSM
101  FTHPSLMGS GVPGHPAIP HPAIVPSSGQ QELQPYDRNL KTOAPEKAK

```

151 EAKKPIKRP LNAFLYKKE MRAKVIACET LKESAINOI LGRRWHLNR
 201 EEOAKYIELA KREROLHMOI YPGMSARDNY GKKRRSRER HOESTDNL
 251 HXSGKRDFAF GTPEKAAP APFLPMVL

!!AA_SEQUENCE 1.0
 PI:S50068 - nonhistone chromosomal protein HMG1-like protein - fruit fly
 (Drosophila melanogaster)
 N:Alternate names: dorsal switch protein
 C:Species: Drosophila melanogaster
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1995
 C:Accession: S50068; S47596
 R:Lehming, N.
 A:Reference number: S50068
 A:Reference number: S50068
 A:Accession: S50068
 A:Molecule type: mRNA
 A:Residues: 1-393 <LEW>
 A:Cross-references: EMBL:U13881; NID:9551089; PID:AA50238.1; PID:9551090
 R:Lehming, N.; Thanos, D.; Brickman, J.M.; Ma, J.; Maniatis, T.; Plasmine, M.
 A:Title: An HMG-like protein that can switch a transcriptional activator to a repressor.
 A:Reference number: S47596; MUID:94352416
 A:Accession: S47596
 A:Molecule type: mRNA
 A:Residues: 178-393 <LEW>
 A:Cross-references: EMBL:U13881
 C:Genetics:
 A:Gene: FLYBase:Dsp1
 A:Cross-references: FLYBase:Fgn011764
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 F:6-47/Region: glutamine-rich
 F:131-161/Region: glutamine-rich
 F:176-253/Domain: HMG box homology <HMG1>
 F:268-342/Domain: HMG box homology <HMG2>

S50068 Length: 393 February 11, 2000 15:51 Type: P Check: 241 ..

1 MEHFHOIQCT IOHYOOQLAA QOOQOVQOOQ LOQHVVYVQ NOQAHOSS
 51 NTAGVGTQO LFTYKASSE PNPAITMAOV VAISNAGTI GVDYLNMAQ
 101 AAAAANPVS QWYSANAG QVDAVTATLT QHQOQOQOQO QOQOQOQOQO
 151 OQOQOQOQO QWINSASPM SRVADAKPR GRMTAVAYV OTCREHKK
 201 HPDEYIAFE FSRKCAERKK TMVDKEKRF HEMAEKDKOR YEAEQNTVP
 251 PKALMDRCK KRKQIKDPNA PKRSLAPFW FCNDEKKNK ALNPEFGVD
 301 IAEELGRKS DVPDYKQRT ESMAERDKAR YEREMTEYTI SGKIASAPS
 351 MQASMQAOAO KALLAARAO QOHOOLEEOH DDDGDGDDDD ENQ

!!AA_SEQUENCE 1.0
 PI:A38095 - T-cell-specific transcription factor 1 splice form A - human
 N:Alternate names: transcription factor TCF-1A
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1995
 C:Accession: A38095; S1645; S65005
 R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Sujiherbulj, R.; Geurts van Kessel, A.; Clevers, H.
 J. Biol. Chem. 267, 8530-8536, 1992
 A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization.
 A:Reference number: A38095; MUID:92235082
 A:Accession: A38095
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-269 <VAV>
 A:Cross-references: GB:X63901

A:Note: authors translated the codon GAT for residue 253 as Ser
 R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.
 EMBO J. 10, 123-132, 1991
 A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.
 A:Reference number: S13449; MUID:91114695
 A:Accession: S1645
 A:Molecule type: mRNA
 A:Residues: 1-269 <WEN>
 A:Cross-references: EMBL:X59869; NID:936785; PID:CAA42526.1; PID:936786
 R:Mayr, K.; Wolff, E.; Clevers, H.; Balhansen, W.G.
 Biochim. Biophys. Acta 1263, 169-172, 1995
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61796; MUID:95367594
 A:Accession: S65005
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 244-269 <MAX>
 C:Genetics:
 A:Intons: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 244/1
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 C:Keywords: alternative splicing; DNA binding; transcription factor
 F:151-226/Domain: HMG box homology <HMG1>

A38095 Length: 269 February 11, 2000 15:51 Type: P Check: 5007 ..

1 MKKEVYSAF NLMMHYPPS GAGOHPOPOP PLRKANOPH GVPOLSYEH
 51 FNSRPIPPAP ADISOKOVR PLQTPDLSGF YSLTSGSMQ LPHYVSWTH
 101 PSLMGSSVP GHPAIPHPA IVPSSGKQEL QPFDRLTKT AESKAEKAK
 151 KPTTKPLNA FMUYKEMRA KVIAECTLKE SAAINOLGR RMWALSREBO
 201 AKYELARKE ROLHMOQYLP WSARDNYGK KRSRREKHOE STTETNWPBE
 251 LKDGNGQEST SMSSSSSPA

!!AA_SEQUENCE 1.0
 PI:B38095 - T-cell-specific transcription factor 1 splice form B - human
 N:Alternate names: transcription factor TCF-1B
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1995
 C:Accession: B38095; S78559; S65006
 R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; Sujiherbulj, R.; Geurts van Kessel, A.; Clevers, H.
 J. Biol. Chem. 267, 8530-8536, 1992
 A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
 A:Reference number: A38095; MUID:92235082
 A:Accession: B38095
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-269 <VAV>
 A:Cross-references: GB:X63901
 R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.
 EMBO J. 10, 123-132, 1991
 A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.
 A:Reference number: S13449; MUID:91114695
 A:Accession: S78559
 A:Molecule type: mRNA
 A:Residues: 1-269 <VAV>
 A:Cross-references: EMBL:X59870; NID:936787; PID:CAA42527.1; PID:936788
 R:Mayr, K.; Wolff, E.; Clevers, H.; Balhansen, W.G.
 Biochim. Biophys. Acta 1263, 169-172, 1995
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61796; MUID:95367594
 A:Accession: S65006
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA

A:Residues: 244-269 <MAY>
 A:Note: only a part of the coding sequence is given
 C:Genetics:
 A:Introns: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 243/3
 N:Alternate names: transcription factor TCF-1C
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1999
 C:Accession: D38095; S61877; S61800
 R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; SuJkerbuijk, R.; Geurts van Kessel, A.; Clevers, H.
 J. Biol. Chem. 267, 8530-8536, 1992
 A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
 A:Reference number: A38095; MUID:92235082
 A:Accession: D38095
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-250 <VAN>
 A:Cross-references: GB:X63901
 R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.
 submitted to the EMBL Data Library, January 1995
 A:Description: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61877
 A:Accession: S61877
 A:Molecule type: mRNA
 A:Residues: 1-250 <MAY>
 A:Cross-references: EMBL:247361; NID:9619881; PIDN:CAA87439.1; PID:9619882
 R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.
 Biochim. Biophys. Acta 1263, 169-172, 1995
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61796; MUID:95367594
 A:Accession: S61800
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 244-250 <MAY>
 A:Cross-references: EMBL:247361
 C:Genetics: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 243/3
 A:Introns: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3; 243/3
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 C:Keywords: alternative splicing; DNA binding; transcription factor
 F:151-226/Domain: HMG box homology <HMG1>
 D38095 Length: 269 February 11, 2000 15:51 Type: P Check: 3201 ..

1 MYKETVSAF NLNHYPPS GAGHPOPOP PLHKANOPH GVPOLSLYEH
 51 FNSPHPTAP ADISOKOVHR PLQTPDLGSG YSLTSGSMQ LPHTVSWFTH
 101 PSLMIGSGVP GHPAIPHPA IVPSPGKOEL OPEDRNLKTO AESKAKEAK
 151 KPTIKKPLNA FMYMKEMRA KVIAECTLIKE SAAINOILGR RWHALSREDO
 201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHQOE STTGGRNMF
 251 GTYEKAAP APFLPMTVL

201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHQOE STTDSNLHYS
 11AA:SEQUENCE 1.0
 P1:J6179 - dorsal switch protein 1 - fruit fly (Drosophila melanogaster)
 N:Alternate names: transcription factor TCF-1C
 C:Species: Homo sapiens (man)
 C>Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 24-Sep-1999
 C:Accession: C38095; S13449; S65007
 R:van de Wetering, M.; Oosterwegel, M.; Holstege, F.; Dooyes, D.; SuJkerbuijk, R.; Geurts van Kessel, A.; Clevers, H.
 J. Biol. Chem. 267, 8530-8536, 1992
 A:Title: The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization
 A:Reference number: A38095; MUID:92235082
 A:Accession: C38095
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-268 <VAN>
 A:Cross-references: GB:X63901
 R:van de Wetering, M.; Oosterwegel, M.; Dooyes, D.; Clevers, H.
 EMBO J. 10, 123-132, 1991
 A:Title: Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box.
 A:Reference number: S13449; MUID:91114695
 A:Accession: S13449
 A:Molecule type: mRNA
 A:Residues: 1-268 <WET>
 A:Cross-references: EMBL:X59871; NID:936789; PIDN:CAA42528.1; PID:936790
 A:Note: the authors describe an additional C-terminal exon of splice form C, the which is not contained in the respective nucleic acid sequence submitted to the EMBL data library and not mentioned in references A38095 and S61796
 R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.
 Biochim. Biophys. Acta 1263, 169-172, 1995
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61796; MUID:95367594
 A:Accession: S65007
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 244-268 <MAY>
 A:Note: only a part of the coding sequence is given
 C:Genetics: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3 244/1
 A:Introns: 32/3; 68/1; 97/2; 137/2; 191/3; 227/3 244/1
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 C:Keywords: alternative splicing; DNA binding; transcription factor
 F:151-226/Domain: HMG box homology <HMG1>
 C38095 Length: 268 February 11, 2000 15:51 Type: P Check: 409 ..

1 MYKETVSAF NLNHYPPS GAGHPOPOP PLHKANOPH GVPOLSLYEH
 51 FNSPHPTAP ADISOKOVHR PLQTPDLGSG YSLTSGSMQ LPHTVSWFTH
 101 PSLMIGSGVP GHPAIPHPA IVPSPGKOEL OPEDRNLKTO AESKAKEAK
 151 KPTIKKPLNA FMYMKEMRA KVIAECTLIKE SAAINOILGR RWHALSREDO
 201 AKYELARKE ROLHMOLYPG WSARDNYGKK KRSREKHQOE STDPGSPFK
 251 CRARFGLNQ TDWCGPCR

11AA:SEQUENCE 1.0
 P1:J6179 - dorsal switch protein 1 - fruit fly (Drosophila melanogaster)
 N:Alternate names: transcription factor TCF-1C
 C:Species: Drosophila melanogaster
 C>Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 24-Sep-1999
 C:Accession: J6179
 R:Canaple, J.; Decoville, M.; Leng, M.; Locker, D.
 Gene 184, 285-290, 1997
 A:Title: The Drosophila DSL gene encoding an HMG 1-like protein: genomic organization, evolutionary conservation and expression.
 A:Reference number: J6179; MUID:97103675
 A:Accession: J6179

A:Molecule type: mRNA
 A:Residues: 1-393 <CAN>
 A:Cross-references: EMBL:X89811; NID:g1150374; PIDN:CAA61938.1; PID:e190037;
 PID:g1150375
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 F:116-253/Domain: HMG box homology <HMG1>
 F:268-342/Domain: HMG box homology <HMG2>

J06179 Length: 393 February 11, 2000 15:51 Type: P Check: 83

1 MEHFHIOIOT IOHYOOILAA OOOOOVOOOO LOOHQVYVQ NOQAONONS
 51 NTAGVGTOQ LFTYKASSF PNPATMAOV VATSNAAGTI GYDRIANMAQ
 101 AAAAAAAYGSG QMWYSAANOG QVDANTAAOL QHQOQOQOQO QOQOQOQOQO
 151 QOQOQOQOQO QVYNSASPM SVKADAKPR GRMTAYAYEV QTCREHKKK
 201 HPDETIVFAE FSRKCAERWK TWVDEKKRF HEMAERKOR YEAEOMNYVP
 251 PKGAVNGRK KKKQIKDPNA PKRSIAEFW FCNDERKKV ALNPEFGVD
 301 IAKELGRKMS DVDEYKOKY ESMERDKAR YEREMTEYKT SKRIAMSAPS
 351 MQAMQMOAQ KALLALAAAO QOHOOLEOH DDDGDGDDDD ENQ

11AA_SEQUENCE 1.0
 P1:S61796 - T-cell-specific transcription factor 1 splice form E - human
 N:Alternate names: transcription factor TCF-1E
 C:Species: Homo sapiens (man)
 C>Date: 28-Oct-1996 #sequence_revision 24-Jul-1998 #text_change 24-Sep-1999
 C:Accession: S61796; S61878
 R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.
 Biochem. Biophys. Acta 1263, 169-172, 1995
 A:Title: The human high mobility group (HMG)-box transcription factor TCF-1:
 novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61796; MUID:95367594
 A:Accession: S61796
 A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 244-366 <MAY>
 A:Note: Only a part of the coding sequence is given
 R:Mayer, K.; Wolff, E.; Clevers, H.; Ballhausen, W.G.
 submitted to the EMBL Data Library, January 1995
 A:Description: The human high mobility group (HMG)-box transcription factor
 TCF-1: novel isoforms due to alternative splicing and usage of a new exon IXA.
 A:Reference number: S61877
 A:Accession: S61878
 A:Molecule type: mRNA
 A:Residues: 1-324, 'RHLPQVPLSASQPGPHRPAACRAHRSNNRLNDRMPSRRHPTGRLQETP',
 <MAY>
 A:Cross-references: EMBL:Z47362; NID:6619883; PIDN:CAA87440.1; PID:g619884
 A:Note: the difference in residues C-terminal of 324 is due to a frameshift
 error
 C:Superfamily: unassigned HMG box proteins; HMG box homology
 C:Keywords: alternative splicing; DNA binding; transcription factor
 F:151-226/Domain: HMG box homology <HMG1>
 S61796 Length: 366 February 11, 2000 15:51 Type: P Check: 5713

1 MKETVYSAF NLMMHYPPS GAGOHPOPO PLHKAPOPH GVPOLSYEH
 51 FNSPHPPAP ADISOKOVNR PLQTPDLGGF YSLTSGSGMQ LPHIVSMETH
 101 PSLMGSGVP GHPAALHPA IYPPSGKQEL QPDRNLKTO ABEKAKKEAK
 151 KPTIKKPLNA FLMYKEMRA KYIACTLKE SAALNOLLGR RMHALSREDO
 201 AKYELARKE ROLHMLYGP WSARDNYGK KRRREKHQE SITDPSSPK
 251 CRAFGINQO TDMCGPCRRK KKCIIYLGE GRCSPPSPD DSALGCGSP
 301 APOSPSYHL LPFPTEILT SPAEAPISP GLSTALSLEPT PGPPAPNST

351 LOSTOVOOE SOROVA

11AA_SEQUENCE 1.0
 P1:J07677 homeobox protein HEX - human
 C:Species: Homo sapiens (man)
 C>Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 24-Sep-1999
 C:Accession: J07677; S26799; S78048; S30231
 R:Hiromas, R.; Radich, J.; Collins, S.
 Biochem. Biophys. Res. Commun. 195, 976-983, 1993
 A:Title: PCR cloning of an orphan homeobox gene (PRH) preferentially expressed
 in myeloid and liver cells.
 A:Reference number: J07677; MUID:93384629
 A:Accession: J07677
 A:Molecule type: DNA
 A:Residues: 1-270 <ACC>
 A:Cross-references: GB:L16499; NID:g292404; PIDN:AA02988.1; PID:g292405
 R:Clompton, M.R.; Bartlett, T.J.; Macgregor, A.D.; Manfioletti, G.; Burattini,
 E.; Giancotti, V.; Goodwin, G.H.
 Nucleic Acids Res. 20, 5661-5667, 1992
 A:Title: Identification of a novel vertebrate homeobox gene expressed in
 haematopoietic cells.
 A:Reference number: S26799; MUID:93087175
 A:Accession: S26799
 A:Molecule type: mRNA
 A:Residues: 1-114, 'L', 116-270 <CHRO>
 A:Cross-references: EMBL:X67235; NID:g32547; PIDN:CAA47661.1; PID:g32548
 R:Bedford, F.K.
 submitted to the EMBL Data Library, February 1993
 A:Reference number: S78048
 A:Accession: S78048
 A:Molecule type: DNA
 A:Residues: 122-270 <BED>
 A:Cross-references: EMBL:Z21533; NID:g32068; PIDN:CAA79730.1; PID:g32069
 R:Bedford, F.K.; Ashworth, A.; Enver, T.; Wiedemann, L.M.
 Nucleic Acids Res. 21, 1245-1249, 1993
 A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and
 conserved between mouse and human.
 A:Reference number: S30230; MUID:93219088
 A:Accession: S30231
 A:Molecule type: DNA
 A:Residues: 136-195 <BEW>
 A:Cross-references: EMBL:Z21533
 C:Comment: This protein is an important regulator of normal hematopoiesis.
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:138-194/Domain: homeobox homology <HOX>
 J07677 Length: 270 February 11, 2000 15:51 Type: P Check: 3795

1 MGYPHGPAA GAVGVPLXAP TPLLOPAHPT PFYEDILGR GRAPAPDAPT
 51 LSPNSSTFS LVSPYRTVPV EPTPLHPAFS HHSAAALAA YGPGGCGPL
 101 YFPRTVNDY THALVRHPL GKPLMSPL QRLPHKRGQ QVRFNSDQTI
 151 ELEKFEYOK YLSEPPERKL AKMLQSERO VKTFQONRRA KWRRLKQEP
 201 QSNKKELES LDSSCDOROD LPSEQNGAS LDSSQCSPP ASQEDLESEI
 251 SEDSDQEVDI EGDKSYFNAG

11AA_SEQUENCE 1.0
 P1:I46089 - thyroid transcription factor-1 - dog
 C:Species: Canis lupus familiaris (dog)
 C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 24-Sep-1999
 C:Accession: I46089; S42104
 R:Van Renterghem, P.H.G.; Dremier, S.; Vassar, G.; Christophe, J.
 Mol. Cell. Endocrinol. 112, 83-93, 1995
 A:Title: Study of TTF1 gene expression in dog thyrocytes in primary culture.
 A:Reference number: I46089; MUID:96034516

A:Accession: 146089
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-371 <VAN>
 A:Cross-references: EMBL:X77910; NID:9457486; PIDN:CA454868.1; PID:9457489
 C:Genetics:
 A:Gene: TTF-1
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:162-218/Domain: homeobox homology <HOX>
 146089 Length: 371 February 11, 2000 15:51 Type: P Check: 4988

1 MMSMRHTTP FSVSDILSPL EESYKKGME GGLGAPLAA YRQQAAPPA
 51 AAMQCHAVGH HGAVTAAYHM TAAGVPOLSH SAVGVCNNGN LGNNSELPY
 101 ODTMRNSAG PGWYGANPD REPALSRMG PASGNNNSGM GGLSLGDVS
 151 KNAAPLPAP RRRRVLFQ AQYELERF KQKYLAP REHLASMIHL
 201 TPTQVIMFQ NHRKMKRQA KDKAAQOQL QDSGGGGGGG GAGCPQOQO
 251 AQOSPRRVA PVLYKDKPC QAGAPAPGAG SLQCHAAQQA AQQAQAQA
 301 AAALSVGSG PGLAHGPHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL
 351 NSSGSDYGM MCTLLYGR W

11AA_SEQUENCE 1.0
 P1:530230 - homeotic protein Hex - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Sep-1999
 C:Accession: S30230
 R:Bedford, F.K.; Ashworth, A.; Emver, T.; Wiedemann, L.M.
 Nucleic Acids Res. 21, 1245-1249, 1993
 A:Title: HEX: a novel homeobox gene expressed during haematopoiesis and conserved between mouse and human
 A:Reference number: S30230; M0ID:93219088
 A:Accession: S30230
 A:Molecule type: DNA
 A:Residues: 1-271 <BED>
 A:Cross-references: EMBL:221524; NID:9288500; PIDN:CA49729.1; PID:9288501
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:139-195/Domain: homeobox homology <HOX>
 S30230 Length: 271 February 11, 2000 15:51 Type: P Check: 8616

1 MGFPHGPAA APAGVPLXA PTPILQPAHP TPYIDILG RGPAPPTPT
 51 TLPSPNSFT SLVSSYRTPV YEPTRVHPAF SHHPAALAA AVGPSGEGP
 101 LYPFRITVND YTHALLRHP LCKPLWSPF LQRPILKRG GGVRESNDQT
 151 VLEKKFETQ KYLSPERKR LAKMLQLSER QVKTWQNR AKWRILKCN
 201 POSNKKDALD SLDTSCGOQ DLPSRQNGKA SLDRSQCSPS PASQEDPDE
 251 ISEDSQEVQ IEDKGFYNA G

11AA_SEQUENCE 1.0
 P1:53724 - thyroid-specific enhancer-binding protein (T/EBP) - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 24-Sep-1999
 C:Accession: S53724
 R:Oguchi, H.; Pan, Y.T.; Kimura, S.
 Biochim. Biophys. Acta 1261, 304-306, 1995
 A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding protein (T/EBP) gene: extensive identity of the deduced amino acid sequence with the human protein.
 A:Reference number: S53724; M0ID:95226463
 A:Accession: S53724

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <OGU>
 A:Cross-references: GB:U19755; NID:9885887; PIDN:AAA86100.1; PID:9885888
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:162-218/Domain: homeobox homology <HOX>
 S53724 Length: 372 February 11, 2000 15:51 Type: P Check: 7622

1 MMSMRHTTP FSVSDILSPL EESYKKGME GGLGAPLAA YRQQAAPPA
 51 AAMQCHAVGH HGAVTAAYHM TAAGVPOLSH SAVGVCNNGN LGNNSELPY
 101 ODTMRNSAG PGWYGANPD REPALSRMG PASGNNNSGM GGLSLGDVS
 151 KNAAPLPAP RRRRVLFQ AQYELERF KQKYLAP REHLASMIHL
 201 TPTQVIMFQ NHRKMKRQA KDKAAQOQL QDSGGGGGGG GAGCPQOQO
 251 AQOSPRRVA PVLYKDKPC QAGAPAPGAG SLQCHAAQQA AQQAQAQA
 301 AAALSVGSG PGLAHGPHQ PGSAGQSPDL AHHAASPAAL QGVVSLSHL
 351 LNSSGSDYGA MCTLLYGR TW

11AA_SEQUENCE 1.0
 P1:A56451 - thyroid-specific enhancer-binding protein (T/EBP) - human
 N:Alternate names: Drosophila NK-2 homolog A; thyroid nuclear factor; thyroid transcription factor 1
 C:Species: Homo sapiens (man)
 C>Date: 27-Oct-1995 #sequence_revision 27-Oct-1995 #text_change 24-Sep-1999
 C:Accession: A56451; S53723; S53725; G02041
 R:Frederick, R.; Clark, J.C.; Shaw-White, J.R.; Stahlman, M.T.; Bouteil, C.J.; Whitsett, J.A. 270, 8108-8114, 1995
 J. Biol. Chem.
 A:Title: Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells
 A:Reference number: A56451; M0ID:95229626
 A:Accession: A56451
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-371 <TKE>
 A:Cross-references: GB:U19816; NID:9767832; PIDN:ANC50125.1; PID:9767833
 R:Salardi, A.; Tassi, V.; de Filippis, G.; Clivatareale, D.
 Biochim. Biophys. Acta 1261, 307-310, 1995
 A:Title: Cloning and sequence analysis of human thyroid transcription factor 1.
 A:Reference number: S53723; M0ID:95226464
 A:Accession: S53723
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-371 <SAI>
 A:Cross-references: EMBL:X82850; NID:9695582; PIDN:CA58053.1; PID:9695583
 R:Oguchi, H.; Pan, Y.T.; Kimura, S.
 Biochim. Biophys. Acta 1261, 304-306, 1995
 A:Title: The complete nucleotide sequence of the mouse thyroid-specific enhancer-binding protein (T/EBP) gene: extensive identity of the deduced amino acid sequence with the human protein.
 A:Reference number: S53724; M0ID:95226463
 A:Accession: S53725
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-241, 'G', 242-371 <OGU>
 R:Hamdan, H.
 Submitted to the EMBL Data Library, August 1995
 A:Reference number: G02041
 A:Accession: G02041
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-225, 'HE', 228-371 <HNM>
 A:Cross-references: EMBL:U03749; NID:91072047; PID:91072048
 C:Genetics:
 A:Gene: GDB:TTF1; NKX2A; TTF-1

A:Cross-references: GDB:132588; OMIM:600635
 A:Map position: 14q13-14q13
 A:Introns: 125/1
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:162-218/Domain: homeobox homology <HOX>
 A56451 Length: 371 February 11, 2000 15:51 Type: P Check: 5439

1 MMSMKHTTP FSVSDILSPL EESYKRYGME GGLGAPLAA YRGOAAPT
 51 AAMQOHAVGH HGAVTAAAYHM TAAGVPOLSH SAVGVCNCGN LGNNSELPY
 101 QDMRNSASG PGWYGANDP RFPALSRFMG PASGMMSGM GGLSLGDS
 151 KNNAPLPSP RRRRVLFQ AOYELERF KOOKYLSAE REHLASMTL
 201 TPPOVKTFQ NRRYKMKRQA KDKAAQOQLQ QDSGGGGGGG GTGCPQOQA
 251 QOQSRRRAV PVLVKDGRPC QAGAPAPGAA SIQGHQOQA QHQAQAQA
 301 AAASVSGG AGLGAHPGHQ PGSAGSPDL AHHAAPAL OGQVSSLSHL
 351 NSSGSDYGTW SCSTLLYGR T W

!!AA_SEQUENCE 1.0
 P1:SI2002 - thyroid nuclear factor 1 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 13-Jan-1995 #sequence_revision 13-Mar-1997 #text_change 24-Sep-1999
 C:Accession: SI2002; PC2252
 R:Gazzzi, S.; Price, M.; de Felice, M.; Damante, G.; Mattel, M.G.; di Lauro, R.
 EMBL J. 9, 3631-3639, 1990
 A:Title: Thyroid nuclear factor 1 (TF-1) contains a homeodomain and displays a novel DNA binding specificity.
 A:Reference number: SI2002; MUID:91006063
 A:Accession: SI2002
 A:Molecule type: mRNA
 A:Residues: 1-372 <CU>
 A:Cross-references: EMBL:X53658; NID:957422; PIDN:CAA37851.1; PID:957423
 A:Note: Met-1 was determined by sequencing DNA
 R:Endo, T.; Ohta, K.; Saito, T.; Haraguchi, K.; Nakazato, M.; Kogai, T.; Onaya, T.
 Biochem. Biophys. Res. Commun. 204, 1358-1363, 1994
 A:Title: Structure of the rat thyroid transcription.
 A:Reference number: PC2252; MUID:95071477
 A:Accession: PC2252
 A:Molecule type: DNA
 A:Residues: 1-35 <END>
 A:Cross-references: DDBJ:D38035
 A:Experimental source: Liver
 C:Comment: This protein activates the genes of thyroid specific protein, thyroglobulin, thyroid peroxidase and thyrotropin receptor.
 C:Genetics:
 A:Gene: TTF-1
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:102-218/Domain: homeobox homology <HOX>
 SI2002 Length: 372 February 11, 2000 15:51 Type: P Check: 6608

1 MMSMKHTTP FSVSDILSPL EESYKRYGME GGLGAPLAA YRGOAAPT
 51 AAMQOHAVGH HGAVTAAAYHM TAAGVPOLSH SAVGVCNCGN LGNNSELPY
 101 QDMRNSASG PGWYGANDP RFPALSRFMG PASGMMSGM GGLSLGDS
 151 KNNAPLPSP RRRRVLFQ AOYELERF KOOKYLSAE REHLASMTL
 201 TPPOVKTFQ NRRYKMKRQA KDKAAQOQLQ QDSGGGGGGG GGACCPQOQ
 251 AOOQSRRRA VVVLVKDKRP CQAGAPAPGA ASIQGHQOQ AOOQAQAQA
 301 AAASVSGG AGLGAHPGH QPSAGSPDL LAHHAAPAL LGQVSSLSHL

351 LNSSGSDYGA MSCSTLLYGR TW

!!AA_SEQUENCE 1.0
 P1:I59234 - octamer binding transcription factor 1 - human
 C:Species: Homo sapiens (man)
 C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 24-Sep-1999
 C:Accession: I59234
 R:Bhargava, A.K.; Li, Z.; Weissman, S.M.
 Proc. Natl. Acad. Sci. U.S.A. 90, 10260-10264, 1993
 A:Title: Differential expression of four members of the POU family of proteins in activated and phorbol 12-myristate 13-acetate-treated Jurkat T cells.
 A:Reference number: I59234; MUID:94052142
 A:Accession: I59234
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-420 <RES>
 A:Cross-references: GB:L20433; NID:9418015; PIDN:AAA65605.1; PID:9418016
 C:Superfamily: unassigned homeobox proteins; homeobox homology; POU domain homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:268-338/Domain: POU domain homology <POU>
 F:357-413/Domain: homeobox homology <HOX>
 I59234 Length: 420 February 11, 2000 15:51 Type: P Check: 9010

1 MMSMKOPH FAMHPTLPEH KYPSSLSSSE AIRRACLPPT PLOSNLFASTL
 51 DETLLARAEV LAADVIVSQ GKSHPEKPA TYHTMNSVPC TSTSTVPLAH
 101 HHHHHHHQA LEPGLDLDHI SSPLALMAG AGAGAAGG GGAHDPPGG
 151 GGGGGGGGG GGGGGGGGG GPGGGGGGG GLLGGSANP HPHMHLGHL
 201 SHPAAANM MPGLPHPL VAAAHNGAA AAAAASAGO VAAASAAAV
 251 VQAAGLASIC DSDTPPRELE AFAERKORR IKLGVQADV GSAALAKLP
 301 GYGLSOSTI CRFSLTSLH NNNIALKPL QAWLEBEGA QRRKMKPEL
 351 ENGGEKRRR TSIAPEKRS LEAYFAVOPR PSSKIAAIA EKLDLKNV
 401 RVWFCNORQ QRRKFSATY

!!AA_SEQUENCE 1.0
 P1:G02321 - thyroid transcription factor 1 - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 24-Sep-1999
 C:Accession: G02321; G02039
 R:Hamdan, H.
 submitted to the EMBL Data Library, December 1995
 A:Reference number: H01047
 A:Accession: G02321
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 1-401 <HA>
 A:Cross-references: EMBL:U43203; NID:91199864; PIDN:AAA89066.1; PID:91199865
 R:Hamdan, H.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: G02039
 A:Accession: G02039
 A>Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-71 <HA2>
 A:Cross-references: EMBL:U33627; NID:91113816; PID:91000129
 C:Genetics:
 A:Gene: TTF-1
 A:Introns: 26/2
 C:Superfamily: unassigned homeobox proteins; homeobox homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:192-248/Domain: homeobox homology <HOX>
 G02321 Length: 401 February 11, 2000 15:51 Type: P Check: 1141

1 MWSGGSGKAR GWEAAGAGRS SPGRLSRRRI MSMSPKHTTP FSVSDLSLPL
 51 EESYKKGME GGLGAPLAA YROGAAPPT AAMOOHAYGH HCAVTAIYIM
 101 TAAGYPOLSH SAVGYCNGN LGMSLELPY QDTRNSASG PGWYCANPDP
 151 RPAISRENG PASGMNSGM GGLSGLDVS KNMAPLDSAP RRRRYLFSQ
 201 AQYELERRF KQKYLSPF REHLASMIH LPTQVKIWFQ NHRYYKKRQA
 251 KDKAAQOOLQ QDSGGGGGG GTCCPOQOQA QOOSPRRYAV PVLKDKKPC
 301 QAGAPAPGAA SLOGHAQQA QHOQAQQA AAATSVGSGG AGLGAPHGQ
 351 PGSAGGSPDL AHNAAPAL QGVVSLSLH NSGSDYGTW SCSTLLYGR
 401 W

11AA_SEQUENCE 1.0
 P1:S78452 - POU domain protein rdc-1 - human
 C:Species: Homo sapiens (man)
 C>Date: 19-Mar-1998 #sequence_revision 19-Mar-1998 #text_change 24-Sep-1999
 C:Accession: S78452; S26063; I38151
 R:Alt, F.W.
 Submitted to the EMBL Data Library, February 1992
 A:Reference number: S78452
 A:Accession: S78452
 A:Molecule type: mRNA
 A:Residues: 1-331 <ALT>
 A:Cross-references: EMBL:X64624; NID:935914; PIDN:CAA5907.1; PID:935915
 R:Collum, R.G.; Fisher, P.E.; Datta, M.; Hells, S.; Thiele, C.; Huebner, K.;
 Croce, C.M.; Israel, M.A.; Thell, T.; Motoy, T.; Depinho, R.; Alt, F.W.
 Nucleic Acids Res. 20, 4919-4925, 1992
 A:Title: A novel POU homeodomain gene specifically expressed in cells of the
 developing mammalian nervous system.
 A:Reference number: I38151; MUID:93027214
 A:Accession: S26063
 A:Molecule type: mRNA
 A:Residues: 174-177, 'S', '179-233', 'K', '235-327' <COL>
 A:Cross-references: EMBL:X64624
 A:Experimental source: Placenta
 C:Superfamily: unassigned homeobox proteins; homeobox homology; POU domain
 homology
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:181-250/Domain: POU domain homology <POU>
 F:268-324/Domain: homeobox homology <HOB>
 S78452 Length: 331 February 11, 2000 15:51 Type: P Check: 3745 ..

1 MNSVPCHTST VPLAHHHHH HHHQALEGD LLDHISPSL ALMAGAGARR
 51 GAGGGGAHDA AGGGGPRGG GPGGGGGGG GGGGAGGGG GPGGGGLGA
 101 SAHPHPHMS LGLSHPPAA AAMNPPSLP HPGIVAAAAA HCAAAAAAAA
 151 AAGVAAASA AAVVGAAGA STCSDSTPR ELEAFSGFK RRIKGYTQA
 201 DVGSALANIK IPGVSLSSQ TICREFSLT SHNMIALKP IIOANLEBAE
 251 GPSEKMPDE LFNGEKKRK RTSIAPEKR SLEAYFAVOP RPSEKIAAI
 301 AEKIDLRNV VRWFNCNOR KOKRMKFSAT Y

11AA_SEQUENCE 1.0
 P1:S78329 - probable phosphoesterase (EC 3.1.1.-) - Synecocystis sp. (strain
 PCC 6803)
 C:Species: Synecocystis sp.
 A:Variety: PCC 6803
 C>Date: 35-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 24-Sep-1999
 C:Accession: S76329
 R:Ranko, T.; Saito, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.;
 Miyajima, N.; Hirosewa, M.; Sugitara, M.; Sasamoto, S.; Kimura, T.; Hosouchi,
 T.; Matsuno, A.; Mureki, A.; Nakazaki, N.; Natsu, K.; Okumura, S.; Shimo, S.;
 Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tanaka, S.
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium
 Synecocystis sp. PCC6803. II. Sequence determination of the entire genome and
 assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S76329
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-416 <KAN>
 A:Cross-references: EMBL:D64000; GB:AB001339; NID:91001484; PIDN:BA010181.1;
 PID:9101832; PID:91001554
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
 1996
 C:Comment: This sequence has motifs characteristic of a variety of
 phosphoesterases.
 C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core
 homology
 C:Keywords: hydrolase
 F:3-95/Domain: phosphoesterase core homology <PEC>
 S76329 Length: 416 February 11, 2000 15:51 Type: P Check: 4655 ..

1 MLKYLHLSDI HUGSGSGSHG INPATGLNTR LEDFQSLRL CIDRALAEPA
 51 DVVLEGGDAP PDATPPYVQ EAFAPERRL ADADIPYVL VGNHDSHSG
 101 SGASLICTYR TLAVPGFVG DRATHLIPT ANGDIQIVTL PWLIRSLILT
 151 RPEEGSLSE AINLLKLRL QPILGEIRS LDPQLTITL AHLMADRAE
 201 GAERTLSGK GFTIPALIN RPEPDYVALG HVHKQNTNP HNPPIVPG
 251 SIERNDEGE KEDRGYIWE ISKGVDMOF CPLPAPFRT IKVDEATD
 301 PGGELMAIA KPAIDNAVR LYOIRSEOL EQIDNKLOE ALKSHSYTI
 351 RPELVQSLAR PRLPGLVGS ALDPLVALKT YLENNKDIKE LMPDMEAE
 401 HLINSEBOW LEAVGD

11AA_SEQUENCE 1.0
 P1:F71508 - probable phosphoesterase (EC 3.1.1.-) CT488 - Chlamydia trachomatis
 (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 24-Sep-1999
 C:Accession: F71508
 R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;
 Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.
 A:Reference number: A71570; MUID:99000809
 A:Accession: F71508
 A:Molecule type: DNA
 A:Residues: 1-244 <ARN>
 A:Cross-references: GB:AE001322; GB:AE001273; NID:93328916; PIDN:AC68088.1;
 PID:93328924
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Comment: This sequence has motifs characteristic of a variety of
 phosphoesterases.
 C:Genetics: CT488
 C:Superfamily: unassigned probable phosphoesterases; phosphoesterase core
 homology
 C:Keywords: hydrolase
 F:2-85/Domain: phosphoesterase core homology <PEC>
 F71508 Length: 244 February 11, 2000 15:51 Type: P Check: 6469 ..

1 MRIFALADH LSLGVPEKTM EVRGEPMGY HOKIEKMRD IYSSDITVL

51 PGDISMANRL EEAQVDFRFL GALPGIKYMI RGNHDYSSA SSARKLVLP
 101 ETLHYSKGY VLLNAHOAIV GVALMDSOI CLHWETQHDG PQRYLFEODD
 151 KIFLEKYGRL ERAKKEIPAS VEDVLVMTYH PVSNDGTGPG RVSNLEMDG
 201 RVSRCLFCHL HKYPRPFGF GNIRGIEYTL VAADYVDICP QVVS

!!AA_SEQUENCE 1.0
 P1:S35782 - serine/threonine protein kinase - bovine herpesvirus 1
 C:Species: bovine herpesvirus 1
 C:Date: 09-Jun-1994 #sequence_revision 12-May-1995 #text_change 24-Sep-1999
 C:Accession: S35782
 R:Audonnet, J.
 Submitted to the EMBL Data Library, June 1993
 A:Reference number: S35782
 A:Accession: S35782
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-467 <AUD>
 A:Cross-references: EMBL:Z23068; NID:9312185; PIDN:CAA80602.1; PID:9312186
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
 C:Keywords: ATP
 F:162-439/Domain: protein kinase homology <KIN>

S35782 Length: 467 February 11, 2000 15:51 Type: P Check: 5295 ..

1 MERAERLAR QRAQLWRSR FACQVAERS GSRLGQSVRG AAAAPARCAA
 51 EGSADLYLAV NNEPEVAPP ARGPDPADG IEGGAAGN EOGGVAAGNE
 101 RRAATGDEKE SASGENESE SESESESE SEADGDMD DDDAGRAGV
 151 TREAEAGAR ALNFRITRL TPGSEGRVE ATGPAPQEH VIKIGASAS
 201 TLAEMLIRT LDHANVVKL AVLEHGLVC AVLARYEDL HTLWKIDRP
 251 MALPTALQVT RAVLOGLAYL HSRRLAHRDV KTEVFLNPG GDVCLDDEGA
 301 AHGPTYEPY YGLAGTLEIN SPELLARARY DCRDWSAG VVAYEMLAVP
 351 RALFDPSPGP QGEDAEASGP PTILGDROCA RQLLRVIRL AVAAEFPSP
 401 PDRLTRFRK RHAATGREPH SPYRCLAVLR LPCDADRLH QMLTFPRAR
 451 PTAELLEHP VFGAASG

!!AA_SEQUENCE 1.0
 P1:C69230 - conserved hypothetical protein MTH973 - Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1999
 C:Accession: C69230
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pochier, B.; Qiu, D.; Spadafora, R.; Viclire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Saffer, H.; Patwell, D.; Prabakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: C69230
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-257 <MTH>
 A:Cross-references: GB:AE000871; GB:AE000666; NID:92622069; PIDN:AAB85469.1; PID:92622072
 A:Experimental source: strain Delta H

A:Gene: MTH973
 A:Start codon: GTG
 C:Superfamily: hypothetical protein M40915

C69230 Length: 257 February 11, 2000 15:51 Type: P Check: 3724 ..

1 MIMWVGIV CGALANLITG FVRDGRVPYK LGFFYDDLE RAENLASWD
 51 GRAVLDVADM LPEYDLYEVA ASPEAVRDYV PEILEAKDV VMSVQALND
 101 PELREMLVEL ASLNDATIHV PGSAIVGLDG LKAASMGNI SVYLITRRPP
 151 RSLGISMDEK KYLYRGASE AVKKEPLNIN VAAALSLACD RDVDVEIIND
 201 PAVDRNVHEV TVRQDFGEFK TITENVCSV NPKTSVMAAY SAIKLKLSIS
 251 ENIHIGT

!!AA_SEQUENCE 1.0
 P1:A40369 - trans-acting transcription regulator rbcR - Chromatium vinosum
 C:Species: Chromatium vinosum
 C:Date: 21-Feb-1992 #sequence_revision 21-Feb-1992 #text_change 24-Sep-1999
 C:Accession: A40369
 R:Viñale, A.M.; Kobayashi, H.; Akazawa, T.; Henikoff, S.
 J. Bacteriol. 173, 5224-5229, 1991
 A:Title: rbcR (sic, rbcR), a gene coding for a member of the LysR family of transcriptional regulators, is located upstream of the expressed set of ribulose 1,5-bisphosphate carboxylase/oxygenase genes in the photosynthetic bacterium Chromatium vinosum
 A:Reference number: A40369; MUID:91317745
 A:Accession: A40369
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-302 <VIA>
 A:Cross-references: GB:M64032; NID:9145018; PIDN:AAA23327.1; PID:9145020
 C:Superfamily: transcription activator LysR-type
 C:Keywords: DNA binding; transcription regulation

A40369 Length: 302 February 11, 2000 15:51 Type: P Check: 4295 ..

1 MHVSRLQRLV FEAVARNISY TRAEELHS QPAVSMQVRQ LEDELISLIF
 51 ERLGRQVILT EAGREVFHYS RAIGOSLREM EEVLESILKV SROSLRIAVA
 101 STVNFAPRL MAIFQSHSG IGLRDVTNR ESTVQMDSN SYDLVLMGVP
 151 PRNVEEADA FMDNPLVIVA PPDHPLAGER AISLRLAE TFMREESGSG
 201 TRQANERFES ERGOTIRHGM QMTRNDVAKO AVRSGLGLSV VSLHTELEL
 251 ERLRVLYDVG EGFPDRRWYV LYRRGKRHS PAAGAFREYV LSEARNMHR
 301 LG

!!AA_SEQUENCE 1.0
 P1:A33960 - cytochrome caa3 oxidase ctaA - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 23-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 24-Sep-1999
 C:Accession: A33960; B65609
 R:Muehler, J.P.; Taber, H.W.
 J. Bacteriol. 171, 4967-4978, 1989
 A:Title: Isolation and sequence of ctaA, a gene required for cytochrome aa-3 biosynthesis and sporulation in Bacillus subtilis.
 A:Reference number: A33960; MUID:89359135
 A:Accession: A33960
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <MUE>
 A:Cross-references: GB:M23915; NID:9142716; PIDN:AAA50254.1; PID:9142717
 A:Note: The authors translated the codon ACT for residue 236 as TTP
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berto, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;

Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, L.F.; Cummings, N.J.; Danile, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Erington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Accession: 139819
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-118 'AT' 20-101 'S', 103-170 'A', 172-188, 'A', 190, 'F', 192-205, 'V', 207-214, 'V', 216-252, 'G', 254-274 'A', 276-279 'T', 281-288, 'T', 290-297 'CQU12'
 A:Cross-references: GB:M94110; NID:9142782; PIDN:AAA2362.1; PID:9142784
 A:Experimental source: strain OF4
 C:Genetics: ctat
 A:Start codon: GTG
 C:Superfamily: cytochrome caa3 oxidase ctat
 A:Accession: A45335
 Length: 297 February 11, 2000 15:51 Type: P Check: 368 ..

1 MKRLKLVSV ITSIGVILV LGGALVTKG SGBSCGATWP LCGEVIPTN
 51 PAETIIIEVS HRIYSGLAGA MIIILAIWAM KQKMHREAK ALSPAAVILI
 101 IFQGLIGAGA AVFGSKAIL ALHFGISAMS LAAYVILITL AFDGRENHTM
 151 AFRVSKGRFY YFEVITTCY GVYSGAYVK HSEATLACDG LPLNGQILFP
 201 GLYGGGANY FHRVAGTILL LFLILIMTW LSRVRRVVL TWAVLSEFL
 251 VVAQFISGIS IVFQNALSV GIHILITISK LFSALSYAM IITRPSH

11AA SEQUENCE 1.0
 P1:G02514 - P2 purinoceptor - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 24-Sep-1999
 C:Accession: G02514
 R:Hammett, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: H01373
 A:Accession: G02514
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-326 'CNA'
 A:Cross-references: EMBL:U52464; NID:91407632; PIDN:AB03572.1; PID:91407633
 C:Genetics: P2Y6
 A:Gene: P2Y6
 C:Superfamily: ATP receptor P2u
 G02514 Length: 326 February 11, 2000 15:51 Type: P Check: 2327 ..

1 MDNGTGALG LPPTCYRE NFKQILIPV YSAVLAAGLP LNICVTIQC
 51 TTRALTRTA VYTLNLAD LLYAGSLPIL IYNAQGDHW PEGDFACRLV
 101 RFLVYANLNG SILFLTCSF QRYLGICHP LPPMKRGGR AAVLCVAVW
 151 LAVTTCLEPT AIFATGIQR NRTVCYDLS PALATHVMPY GMALTVIGFL
 201 LPFAALLACY CLACRLCRQ DGPAPPAOE RRGKAARVAV VAAAPAFISF
 251 LPFHITKAY LAVSTPGVP CTVLEAFAAA YKGTREFAA NSVLDEILEY
 301 FTQKKFRRRP HELLOKIFAK WQROGR

11AA SEQUENCE 1.0
 P1:JC4800 - P2Y6 receptor - human
 C:Species: Homo sapiens (man)
 C>Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 04-Sep-1998
 C:Accession: JC4800
 R:Communi, D.; Parmentier, M.; Boeynaems, J.M.
 Biochem. Biophys. Res Commun. 222, 303-308, 1996
 A:Title: Cloning, functional expression and tissue distribution of the human P2Y6 receptor.
 A:Reference number: JC4800; MUID:96222498
 A:Accession: JC4800
 A:Molecule type: mRNA
 A:Residues: 1-326 'CNA'
 A:Note: sequence extracted from NCBI backbone (NCBI:P121483); this sequence is the sequence shown in Fig. 2

Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, L.F.; Cummings, N.J.; Danile, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entlan, K.D.; Erington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Accession: 139819
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-118 'AT' 20-101 'S', 103-170 'A', 172-188, 'A', 190, 'F', 192-205, 'V', 207-214, 'V', 216-252, 'G', 254-274 'A', 276-279 'T', 281-288, 'T', 290-297 'CQU12'
 A:Cross-references: GB:M94110; NID:9142782; PIDN:AAA2362.1; PID:9142784
 A:Experimental source: strain OF4
 C:Genetics: ctat
 A:Start codon: GTG
 C:Superfamily: cytochrome caa3 oxidase ctat
 A:Accession: A45335
 Length: 297 February 11, 2000 15:51 Type: P Check: 368 ..

1 MKRLKLVSV ITSIGVILV LGGALVTKG SGBSCGATWP LCGEVIPTN
 51 PAETIIIEVS HRIYSGLAGA MIIILAIWAM KQKMHREAK ALSPAAVILI
 101 IFQGLIGAGA AVFGSKAIL ALHFGISAMS LAAYVILITL AFDGRENHTM
 151 AFRVSKGRFY YFEVITTCY GVYSGAYVK HSEATLACDG LPLNGQILFP
 201 GLYGGGANY FHRVAGTILL LFLILIMTW LSRVRRVVL TWAVLSEFL
 251 VVAQFISGIS IVFQNALSV GIHILITISK LFSALSYAM IITRPSH

11AA SEQUENCE 1.0
 P1:G02514 - P2 purinoceptor - human
 C:Species: Homo sapiens (man)
 C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 24-Sep-1999
 C:Accession: G02514
 R:Hammett, F.; Southey, M.C.; Somers, G.R.; Hutchins, A.M.; Venter, D.J.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: H01373
 A:Accession: G02514
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-326 'CNA'
 A:Cross-references: EMBL:U52464; NID:91407632; PIDN:AB03572.1; PID:91407633
 C:Genetics: P2Y6
 A:Gene: P2Y6
 C:Superfamily: ATP receptor P2u
 G02514 Length: 326 February 11, 2000 15:51 Type: P Check: 2327 ..

1 MDNGTGALG LPPTCYRE NFKQILIPV YSAVLAAGLP LNICVTIQC
 51 TTRALTRTA VYTLNLAD LLYAGSLPIL IYNAQGDHW PEGDFACRLV
 101 RFLVYANLNG SILFLTCSF QRYLGICHP LPPMKRGGR AAVLCVAVW
 151 LAVTTCLEPT AIFATGIQR NRTVCYDLS PALATHVMPY GMALTVIGFL
 201 LPFAALLACY CLACRLCRQ DGPAPPAOE RRGKAARVAV VAAAPAFISF
 251 LPFHITKAY LAVSTPGVP CTVLEAFAAA YKGTREFAA NSVLDEILEY
 301 FTQKKFRRRP HELLOKIFAK WQROGR

11AA SEQUENCE 1.0
 P1:JC4800 - P2Y6 receptor - human
 C:Species: Homo sapiens (man)
 C>Date: 15-Oct-1995 #sequence_revision 16-Aug-1996 #text_change 04-Sep-1998
 C:Accession: JC4800
 R:Communi, D.; Parmentier, M.; Boeynaems, J.M.
 Biochem. Biophys. Res Commun. 222, 303-308, 1996
 A:Title: Cloning, functional expression and tissue distribution of the human P2Y6 receptor.
 A:Reference number: JC4800; MUID:96222498
 A:Accession: JC4800
 A:Molecule type: mRNA
 A:Residues: 1-326 'CNA'
 A:Note: sequence extracted from NCBI backbone (NCBI:P121483); this sequence is the sequence shown in Fig. 2

A:Cross-references: EMBL:X57058
 A:Experimental source: Placenta
 C:Superfamily: ATP receptor P2u
 C:Keywords: glycoprotein; placenta; receptor; transmembrane protein
 F:25-52/Domain: transmembrane #status predicted <TM1>
 F:63-86/Domain: transmembrane #status predicted <TM2>
 F:104-122/Domain: transmembrane #status predicted <TM3>
 F:143-167/Domain: transmembrane #status predicted <TM4>
 F:193-216/Domain: transmembrane #status predicted <TM5>
 F:241-264/Domain: transmembrane #status predicted <TM6>
 F:283-305/Domain: transmembrane #status predicted <TM7>
 F:5.173/Binding site: carbohydrate (Asn) (covalent) #status predicted

Jc4800 Length: 328 February 11, 2000 15:51 Type: P Check: 7545 ..

1 MEMONGCOA LGUPPTCY RENKOLLP PYSAVIAAG LPLNCVITO
 51 ICTRRALTR TAVYTINLAL ADLXACSLP LLTYNAQSD HMPEDFACR
 101 LVRELFYANL HGSILEFTCI SFORYLGICH PLAPWHRKG RRAANLVCA
 151 VWIAVTTQCL PRAIFATGI QNRNVCYDL SPALATHYV PYGMALTVIG
 201 FLUPFAALLA CYCLACRLC RODGPAPVA QERRGRAAM AVVAAPAI
 251 SLEPHITKT AYLAARSTPG VPCVLEAFA AAYKGRPPA SANSVLDPIL
 301 FYFOKKRR RPHELOKLT AKWROGR

!!AA_SEQUENCE 1.0
 P1:P64471 - Hypothetical protein homolog MJ1375 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 04-Sep-1998
 C:Accession: F64471
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.C.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Melman, J.E.; Fuhrman, J.L.; Nguyen, D.; Venter, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hrust, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.
 A:Reference number: A64300; MUID:96337999
 A:Accession: F64471
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-415 <RUL>
 A:Cross-references: GB:U67577; GB:L77117; NID:g1592013; PID:g1592020;
 TIGR:MJ1375; PID:g1511391
 C:Genetics:
 A:Map position: REV1324199-1322952
 C:Superfamily: hypothetical protein MJ1375

F64471 Length: 415 February 11, 2000 15:51 Type: P Check: 3696 ..

1 MSLCKDSIYI LMSNLYSKM AYLFYITAF LLGTEAFGI KGLMPIDTL
 51 TIFSSGIPP AIAKLAEEK EVDINKYIPI LYMILLSVY GELIPYIKY
 101 ILGGHYLNP NLTYAVGLC VVASTVAFS RGILOGLKN KYLSLTWIVE
 151 YTAKVILVEI LLLYLIGIFS LLISLAVLV GGIFFYLLY KALKGFEFK
 201 KLIDIKNTK NIFSNNLDI LRSIPALT SSSYRFGDI DNIVMSING
 251 GFMGSIYGS SXISGIMWF ASANSIPLP RISKTDLSI LKEGIQNTI
 301 FSSIFVIGCL FPEEPLIAF FKTANPEIL CLRIALSSS FMSYTLISS

351 ALOGIGYAKI SYTILFELV LNTILNLYV NAYIVGSL ATLITISIVF
 401 LIVERFALRI KKHII

!!AA_SEQUENCE 1.0
 P1:A47201 - Bombesinlike peptide - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 04-Sep-1998
 C:Accession: A47201
 R:Wechsberger, C.; Krell, G.; Richter, K.
 Proc. Natl. Acad. Sci. U.S.A. 89, 9819-9822, 1992
 A:Title: Isolation and sequence of a cDNA encoding the precursor of a bombesinlike peptide from brain and early embryos of Xenopus laevis.
 A:Reference number: A47201; MUID:93028554
 A:Accession: A47201
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-120 <REC>
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBI:115857, NCBI:115858)
 C:Superfamily: ranatensis

A47201 Length: 120 February 11, 2000 15:51 Type: P Check: 1219 ..

1 MSAPFLRML PLRFLHLL LSFPLFCM ESEDAKRIE KIRGNOMAI
 51 GFMGKRSIQ DYNPEODM DSEDFRPII EMIACTFROE PIRALSPKQ
 101 DEIQWMLKRI MDDYIKITQK

!!AA_SEQUENCE 1.0
 P1:E65139 - Hypothetical protein b3434 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C>Date: 13-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Sep-1999
 C:Accession: E65139
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65139
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-197 <BLAT>
 A:Cross-references: GB:AE000420; GB:U00096; NID:g1789840; PIDN:AACT6459.1;
 PID:g1789842; UMG:p.b3434
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: conserved hypothetical protein MJ1677

E65139 Length: 197 February 11, 2000 15:51 Type: P Check: 1678 ..

1 MNEIISAVAL LILMDPLGN LPIFMSVLKH TPKRRRAIM VRELLIALV
 51 MVLFRAGER ILAFLSLAE TVTSISGII FLIAKMFIP SASNSSGGLP
 101 AGEPEPIVPL AIPVAGPTI LATLMLSHQ YQNGHGLVI ALLAMGTF
 151 VILLOSLFL RLKGRGVNA LERLMGLIV MAAQMFIDG IRMMWKG

!!AA_SEQUENCE 1.0
 P1:S48737 - kynurenine aminotransferase - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 26-Aug-1999
 C:Accession: S48737
 R:Mosca, M.; Corzì, L.; Breton, J.; Speciale, C.; Okuno, E.; Schwarcz, R.; Benatti, L.
 FEBS Lett. 353, 21-24, 1994
 A:Title: Molecular cloning of rat kynurenine aminotransferase: identity with glutamine transaminase K
 A:Reference number: S48737; MUID:95010756
 A:Accession: S48737
 A:Status: preliminary

A:Molecule type: mRNA
 A:Residues: 1-423 <MOS>
 A:Cross-references: GB:S74029; NID:9688257; PID:9688258
 C:Superfamily: aspartate transaminase

S48737 Length: 423 February 11, 2000 15:51 Type: P Check: 6122

1 MTRKLRRL DGDONLWVE FGKLTKEYDV VNLGGFDPF SPPDFATQAF
 51 QQAATSGFEML NQYTRAFGP PLTVNLASF GKLLGQEMP LTNLVTVGA
 101 YGALFTFAQA LVDEGEVIL MEPAFDCEP MTMAAGCPV FVTLKPSAP
 151 KGRGASNDM QDPAELASK FTPTKVLV NTPNNPLGKV FSRMELEVA
 201 NLCOQHDVVC ISDEVYQWLV YDGHQVISA SLPGMMDRL TIGSAGKSES
 251 ATGKRVGWM GPDNIMKHLR TVHONSIFHC PTOQAANAQ CFERQOQHG
 301 QPSSTFLOLP QAMELNRDM IRSLOSGLK LMSQSYPL IADISDFKSK
 351 MPDIPGADE PYDRRAKMA IKMGLVGIP VSTFFSRPHO KDFHVIIFC
 401 FVKDKATLQA MDERLRKKE LQP

11AA_SEQUENCE 1.0
 P1:JQ2268 - O-methyltransferase (EC 2.1.1.-) - maize
 C:Species: Zea mays (maize)
 C>Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 29-Sep-1999
 C:Accession: JQ2268
 R:Held, B.M.; Wang, H.; John, I.; Wurtele, E.S.; Colbert, J.T.
 Plant Physiol. 102, 1001-1008, 1993
 A:Title: An mRNA putatively coding for an O-methyltransferase accumulates preferentially in maize roots and is located predominantly in the region of the endodermis.
 A:Reference number: JQ2268; MUID:94105316
 A:Accession: JQ2268
 A:Molecule type: mRNA
 A:Residues: 1-364 <HEL>
 A:Cross-references: GB:114063; NID:9404069; PID:9404070
 A:Experimental source: root, cv. NKH31
 C:Superfamily: O-methyltransferase
 C:Keywords: methyltransferase

JQ2268 Length: 364 February 11, 2000 15:51 Type: P Check: 6645

1 MELSPNNSTD OSLDAQLEL WTTTFPMKS MALSAIHLR IADAIHLHG
 51 AASLSQILSK VHLHPSRVSS LRRLMRYLFT TNVFGTQPLG GGSDDSEFY
 101 YLTPVSRLL TGSOSQOLAQ TPIANVLDP TIVSPFELG AMFQHELPP
 151 CLEKTHGRG IMELTKDDAT FDALVNDGLA SDSQILYDVA IKQSAEYFG
 201 ISSLVGVGGG IGAAQAALSK APFHVCSVL DLAVVAKAP THTDVQFLAG
 251 DMESIPPAD AVLLKSVLHD WDHDQVKIL KNCKRAIPR EAGKVIIN
 301 MVVGAPEPM KHEKQALFD VYIMFNGME RDEDEMSKIF SEAGSYDRI
 351 IPIVGRSII EYYP

11AA_SEQUENCE 1.0
 P1:G70165 - aldose reductase homolog - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 26-Aug-1999
 C:Accession: G70165
 R:Fraser, C.M.; Castjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Venter, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, L.; Matthews, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.;

Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: G70165
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-315 <HEL>
 A:Cross-references: GB:AE001154; GB:AE000783; NID:92688431; PID:92688432;
 IIGR:B80528
 C:Experimental source: strain B31
 C:Superfamily: conserved hypothetical protein YP1088w

G70165 Length: 315 February 11, 2000 15:51 Type: P Check: 3950

1 MNMLDKINT YSKLLIGSQ FGGYFKOVE KETAKKILK AYDQIRNID
 51 TARAYNGIS EKIIGELIER DPTIRENLI ASKCYMEIS EYENFNEISL
 101 KILKTDYIDI YYIHWKADF DLRIPIASEL EMKVKGRIRY VGSNFEIHS
 151 MESIKRVCKI DVNOIGYNPL FRNKEKDVIP YCEDNNIAT SYSTIAQGL
 201 SKANIKDKNK FNDIREKLI LFKKEIWEYT LKTINKLEI AKINMLTLE
 251 LYSWLKRTK LSGRIVFSK ENYVESNVNS FKAELNDKY EETISILDNF
 301 NHQTNFNL FNKKI

11AA_SEQUENCE 1.0
 P1:G69076 - molybdopter in biosynthesis protein Moeb homolog - Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Sep-1999
 C:Accession: G69076
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Basilladen, K.; Blakey, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keegle, P.; Lum, M.; Pothier, B.; Qiu, D.; Spadefora, R.; Vicaire, R.; Wang, Y.; Wietzbowski, J.; Gibso, R.; Jiwani, N.; Caruso, A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.; McDonnell, S.; Shiner, G.; Goyal, A.; Pietrovski, S.; Church, G.M.; Daniels, C.U.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: G69076
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-251 <MTH>
 A:Cross-references: GB:AE000917; GB:AE000666; NID:92622689; PID:92622694
 A:Experimental source: strain Delta H
 C:Genetics: MTH1571
 C:Superfamily: molybdopter in biosynthesis protein moeb

G69076 Length: 251 February 11, 2000 15:51 Type: P Check: 4836

1 MPEYEGMAY WEMVSROMGL LSRADQLKX DSTVSVICG GIGGAEMVL
 51 ARMVGSLRI IDSDVFDVSN INQLMSSFR DLRIKXVYA AERIRIVNPE
 101 SRKVYHEIF DEENAEIIP GSDAVVDALD NITSKVIAR RCHSEGIPFI
 151 HGAHSGMGQ VTFVREGSPS YEELFRLPSV GELTADYKA KLRDLSKTP
 201 PVIGPVANIT GCLQAAEYIK LITGRGDVIA APRMKFDLF QGEFFKIVEL
 251 S

11AA_SEQUENCE 1.0
 P1:S56490 - hypothetical transcription regulator pepa-gntv intergenic region -

Escherichia coli
 N:Alternate names: hypothetical protein f332
 C:Species: Escherichia coli
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 29-Sep-1999
 C:Accession: S56490; C65239
 R:Burkhardt, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
 Nucleic Acids Res. 23, 2105-2119, 1995
 A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region
 from 92.8 through 100 minutes.
 A:Reference number: S56314; MUID:95334362
 A:Accession: S56490
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-332 <BUR>
 A:Cross-references: EMBL:U14003; NID:91263172; PID:AA97161.1; PID:9537106
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
 J.; Davis, N.W.; Karpatick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:9742617
 A:Accession: C65239
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-332 <BLAT>
 A:Cross-references: GB:AE000497; GB:U00096; NID:91790711; PID:AACT7221.1;
 PID:9190715; UMEP:D4264
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: yjg5
 C:Superfamily: gntR protein
 S56490 Length: 332 February 11, 2000 15:51 Type: P Check: 2843 ..

1 MRNRHLEOD IATLAVGTVM TVSRYSRSPK KVARETGERI AKIMEINYI
 51 PNRAFGMLN AOSYTLGILI PSFQNLFPAD ILAGIESVTS ENHYQILLIN
 101 YNDRDSEEE SVINILSYNI DGIILSEKYH TIRTVKRLS AIIIPVELMD
 151 VQGERLMEV GFDRNROAFD MVTCLERKVR RHKILYIGSK DDTREQRKC
 201 GYCDAMMLN LSPLRNPNRA ISSIHLGMQL MRDALSANPD LQGVCTNDD
 251 IANGALLCR ENRLAVPEQI SIAGHGLEI GRQMPSLAS VTPRPDIGN
 301 MAQMLLSKI KNNHNHNTV DLGQIYHGN TL

11AA SEQUENCE 1.0
 P1:H69823 - conserved hypothetical protein yHt - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
 C:Accession: H69823
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo,
 V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.;
 Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Bron, S.; Broutier, S.;
 Busch, C.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,
 J.J.; Conerton, I.F.; Cummings, N.J.; Daniel, R.A.; Deniset, F.; Devine, K.M.;
 Dueselherdt, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.;
 Fiebel, C.; Ferrati, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi,
 A.; Gallier, N.; Gilm, S.Y.; Glasner, J.; Goffeau, A.; Goldschmidt, E.J.; Grand,
 G.; Guillep, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Hentut, A.;
 Hilbert, H.; Holsappel, S.; Hosono, S.; Hult, M.F.; Itaya, M.; Jones, L.;
 Joritz, B.; Karamata, D.; Kasahara, Y.; Kjaer-Blanchard, M.; Klein, C.;
 Kobayashi, Y.; Koster, S.; Kohnstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;
 Laplante, A.; Lardinois, S.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
 S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;
 Nakai, S.; Noback, M.; Noone, M.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega,

B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
 A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds,
 S.; Rieger, M.; Rivolta, C.; Kocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato,
 T.; Scanlon, E.
 A:Authors: Schlach, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska,
 A.; Serot, S.J.; Serrot, P.; Shin, B.S.; Solido, B.; Sorokin, A.; Taccini, E.;
 Takagi, I.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka,
 T.; Terpe, P.; Toth, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.;
 Vannier, F.; Vassarotti, A.; Vairi, A.; Wambut, R.; Wedler, E.; Wedler, H.;
 Weitzenger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto,
 K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Dancho, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: H69823
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-302 <KUN>
 A:Cross-references: GB:Z99108; GB:AL009126; NID:92633055; PID:CA12749.1;
 PID:9182910; PID:92633244
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yHt
 C:Superfamily: conserved hypothetical protein H10176
 H69823 Length: 302 February 11, 2000 15:51 Type: P Check: 1929 ..

1 MNQKRGLEI LINKKQDQW LSVLKTALK ASKPYIDDM SHOOIVNHE
 51 SVLNMIYK GDRVFDLQ SEASSVPEY GELDILEDN HMLINKPAG
 101 IATHNEDQO TGLANLAI HQINGETCK VRHVRLDQD TSGAIVFAK
 151 RLARHLDQO LEKTLKRTY TAIAGKLT KGTINPPIG RDSHPTRR
 201 VSPGQIVAT HFVYMASNAK ERLSVELEL ETGRTHOIRV HIASLGHPT
 251 GDSLYGGSK LNRQALHAN KVQAVHPITD ELIYAEPFP ADMKNLCRTY
 301 FS

11AA SEQUENCE 1.0
 P1:A64223 - probable pseudouridine synthase - Mycoplasma genitalium (SGC3)
 C:Species: Mycoplasma genitalium
 C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
 C:Accession: A64223; T09718
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.;
 Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.;
 Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhmann, J.L.;
 Nguyen, D.; Uterback, T.R.; Sander, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb,
 J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Luetter, T.S.; Peterson, S.N.;
 Smith, H.O.; Hutchison III, C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; MUID:96026346
 A:Accession: A64223
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1308 <TIG>
 A:Cross-references: GB:U039698; GB:L43967; NID:91045891; PID:91045895; TIGR:MG209
 A:Experimental source: strain G-37
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.;
 Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.G.; Kelley, J.M.;
 Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhmann, J.L.;
 Nguyen, D.T.; Uterback, T.; Sander, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb,
 J.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Luetter, T.S.; Peterson, S.N.; Smith,
 H.O.; Venter, J.C.
 Submitted to the EMBL Data Library, October 1998
 A:Reference number: Z16818
 A:Accession: T09718
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-308 <FRA>
 A:Cross-references: EMBL:U39700; NID:93844800; PID:93844807
 A:Experimental source: isolate G37
 C:Genetics: K6209
 A:Gene: M6209
 A:Genetic code: SGC3
 C:Superfamily: conserved hypothetical protein HI0176
 A64223 Length: 308 February 11, 2000 15:51 Type: P Check: 590 ..

1 MROCFVYVTT KRDSLSL LNSRYKVKV LNMGIKVN EKLTFKNSLI
 51 VARDVYKVE IHDETSDFI TSVEPVNKL EYLFEDKDL VINKPSGLLT
 101 HPTTFNEKAS LLAACIFHNN KNPIYLVHL DDTSGALIV CKNOASLNL
 151 QNOLONETLK RYVALVHP FNAITGSINA PLARVNNKV MEKIAOTAKA
 201 KQATTKKVI NONKKAALS LEILTGRTQ INVHLKTIQH PYINDPLTGI
 251 KSEKDSYQG FLHNRICFI HPTLNKPMDF HAPLEPKFST KLSLSLSLT
 301 DPLVLFK

!!AA_SEQUENCE 1.0
 PI:573869 - hypothetical protein ycec - Mycoplasma pneumoniae (ATCC 29342)
 N:Alternate names: hypothetical protein HI0_crf309
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 26-Aug-1999
 C:Accession: S73869
 R:Himmelfrich, R., Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Hermann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
 A:Reference number: S73327; MUID:9710585
 A:Accession: S73869
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-309 <HIM>
 A:Cross-references: EMBL:AE000053; GB:U00089; NID:q1674236; PID:q1674240
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Gene: ycec
 A:Genetic code: SGC3
 C:Superfamily: conserved hypothetical protein HI0176
 S73869 Length: 309 February 11, 2000 15:51 Type: P Check: 6533 ..

1 MEGTFEVTA QRLDTPLATL LNSRYKVKV LIVDGLSVN GKRTKNGWL
 51 VQEDRVHVN WSELEFEKVP VEQVPYDFPL DILYEDQIM VVKNPGLIS
 101 HPTSFNESES LLGAALFHCN HOPVFLVHL DDTSGVIML AKNOSLTLH
 151 QKQLOQRVAK RYVALVHP LDSLSTISA PLEKGVNKKV MMRVGNSSNK
 201 AKNAFTKTV LNQNKAKALI KCBLLGRTH QIRVHLQRTK HPTVNDPLVG
 251 LKSDQATEYG QYLAQQLSF IHTLNKEMG FEAQDKRFS DKLDNINLKI
 301 ANSLVLFQ

!!AA_SEQUENCE 1.0
 PI:564026 - hypothetical protein HI1338 - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
 C:Accession: B64026
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirmess, E.F.; Kertlage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.,

McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weisman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Uterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhman, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630
 A:Accession: B64026
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-164 <TIGR>
 A:Cross-references: GB:U32813; GB:I42023; NID:91574796; PIDN:AAC22985.1; PID:91574807; TIGR:HI1338
 C:Superfamily: hypothetical protein HI1338
 B64026 Length: 164 February 11, 2000 15:51 Type: P Check: 2974 ..

1 MNIFIMHGE AEVANSRDKA RHLTVGSKQ AFLQGQWLKQ HSTLVINSL
 51 DRIVSPYR AOETFEQVNO APDLEENRF EIWEGITPYG HAHVIDYLE
 101 VLKDEGVKSV LIVSHLPVVG EIVAEVYGRK NPISFPATI AOLMDGNKS
 151 EILMHQASPV IYLK

!!AA_SEQUENCE 1.0
 PI:B65007 - hypothetical protein b2340 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
 C:Accession: B65007
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: B65007
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-161 <BLAT>
 A:Cross-references: GB:AE000322; GB:U00096; NID:91788672; PIDN:AAC75400.1; PID:91788681; UMG:P:D2340
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: hypothetical protein HI1338
 B65007 Length: 161 February 11, 2000 15:51 Type: P Check: 9360 ..

1 MGVFIMRHD AALDAASDV RELTNGCDE SRLANWKG QKVELERYLV
 51 SPFLAREQTL EEWGDCINLP SSAEVLPELT PCGDVGIVSA YLOALINQGV
 101 ASYLVASHLP LVGYLVNELC PETPPEMTT SATASVTLDE SGNTEFNQM
 151 SPONLMKAKA I

!!AA_SEQUENCE 1.0
 PI:C64882 - probable transcription regulator ycjZ - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
 C:Accession: C64882
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: C64882
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-299 <BLAT>
A:Cross-references: GB:AE000231; GB:U00096; NID:g1787388; PIDN:AAC74410.1;
PID:g1787589; UMGF:b1328
A:Experimental source: strain K-12, substrain MG1655
A:Gene: yjc12
C:Superfamily: hypothetical protein b1328
C:Keywords: DNA binding; transcription regulation
F:20-50/Region: regulatory protein lysR motif

C64882 Length: 299 February 11, 2000 15:51 Type: P Check: 51

1 MKREIADLM AFVVAEERS FTRAARLSM AQSALSQIVR RIERLGLRL
51 LRTTRSVAP TEAGEHLLSV LGPMLHDIDS AMASLSDION RPSGTRITTT
101 VEHAKTILL PAMRTFLKSH PEIDIOITLD YGLTDVYSER FDAGVRLGGE
151 MDKMIATRI GPDIPMAIVG SPDYFSRREV PTYSQILDH QAINLILPTS
201 GTANRWMLIR GGREVRYRME GQLLNTITDL IDDAIDGCG LAYLPDYVE
251 RAIEKELKIR VLDKFTPDLP GYHLYPPHRR HAGSAFSLFT DRLKTKGAV

!!AA_SEQUENCE 1.0
P1:E72058 - Probable phosphatase/kinase - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 26-Aug-1999
C:Accession: E72058
R:Kilman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
Grimwood, J.; Davis, R.W.; Stephens, R.S.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; NID:9220606
A:Accession: E72058
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-202 <RNA>
A:Cross-references: GB:AE001645; GB:AE001363; NID:g4376896; PID:g4376907
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: yACE
C:Superfamily: conserved hypothetical protein YDR196C
E72058 Length: 202 February 11, 2000 15:51 Type: P Check: 4148

1 MRLKRVSTI GDLSSGKTEA CQVFOELGAY VVSADDEISHS FLIPHTRIGR
51 RVIDLGSUV VVDGAFDAQA IAAKVFYNSV LLOGLEALIH PEVCRIITEQ
101 YHOSIODGNY PLFVAEVELL YEIHVAKWPD SVILWAMND IRRERFKKT
151 GRSEDFDR CSRLNVEEK LAQADVVEN NQTKRELKOK IEEYFALKG
201 AL

!!AA_SEQUENCE 1.0
P1:B71509 - Probable phosphatase/kinase - Chlamydia trachomatis (serotype D,
strain UM3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 26-Aug-1999
C:Accession: B71509
R:Stephens, R.S.; Kilman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;
Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.
A:Reference number: A71570; NID:99000809
A:Accession: B71509
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-202 <RNA>
A:Cross-references: GB:AE001322; GB:AE001273; NID:g3328916; PID:g3328928

A:Experimental source: serotype D, strain UM-3/Cx
C:Genetics:
A:Gene: yACE
C:Superfamily: conserved hypothetical protein YDR196C
B71509 Length: 202 February 11, 2000 15:51 Type: P Check: 4232

1 MDLKIYSTV GDSGSKTEA CQVFDLGA VISAQVSHS FLVPTSVQ
51 RIIDLGPET IENTLSRKA IAEKVGNRD LLSLEBILH PEVCRVEEK
101 YAHVQOEKY PLFIAEPPL YEIOYADWFD QVILISADTG IKRERFLKKT
151 GGSIDSFDLR CARFSSLEEK ILRADVIEN NGTKKEFRK VQCFKALCG
201 TI

!!AA_SEQUENCE 1.0
P1:D64151 - hypothetical protein H10404 - Haemophilus influenzae (strain Rd
KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: D64151
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
McKenney, K.; Sutton, G.; Fitzhugh, J.F.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips,
C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;
Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
Furumasa, J.L.; Geoghagan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
Venter, J.C.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.
A:Reference number: A64000; NID:93350630
A:Accession: D64151
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-430 <TRIG>
A:Cross-references: GB:U32723; GB:L42023; NID:g1573363; PIDN:AAC22063.1;
PID:g1573375; TIGR:H10404
A:Note: Best homolog was a hypothetical protein from Bacillus subtilis
C:Superfamily: hypothetical protein H10404
D64151 Length: 430 February 11, 2000 15:51 Type: P Check: 1410

1 MDLSDIEKQ LOKATAQAFI IALSGGLDST VLSLFAKLC QKPHPLPPLS
51 IRAIHGHGL SPNADSWAKH CODLQDQFQI PLIERVOYD KTGICAGAR
101 EARYQAIKKY LQTOEMLVTA HHLNDQTEF FLAKRSGSL KGLGAMQOS
151 VLFQPIILP ILGFTFQLE NYAOKEXLMW ITDSNEDNR YDRNFRNFI
201 LPELERNVAH FDLAVQRSNQ HCFEQOOLIN DLSEITTEH COIKNOFKIC
251 QFQYSLAKO TALLRWLAE NQLEMPKRO LFOILINDIV AKEANPOFO
301 LVNKVIRIYQ DSIXLKTPFS DITKCTLKE QNTLNLPDL GNLTVOENEH
351 NLITWODYS VLEKTNLPI SIRQYSGKV KHPKPRRD IKKIWOELGV
401 PPMERNRPL IFYGNELKSA VGFFRVLKSS

!!AA_SEQUENCE 1.0
P1:S75967 - hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C>Date: 23-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 29-Sep-1999
C:Accession: S75967
R:Kaneh, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asaniz, E.; Nakamura, Y.;
Miyajima, N.; Hiroseawa, M.; Sugitara, M.; Sasamoto, S.; Kimura, T.; Hosouchi,

T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3, 109-116, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S75967
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-271 <KAN>
 A:Cross-references: EMBL:D64006; GB:AB001339; NID:g1001291; PIDN:BA010814.1; PID:d011465; PID:g1001327
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics: GTG
 A:Start codon: conserved hypothetical protein H1191
 C:Superfamily: conserved hypothetical protein H1191
 S75967 Length: 271 February 11, 2000 15:51 Type: P Check: 3587 ..

1 MEKCDKNPF GGVREOTKTL PTCGPGSSND HNRLTAESM TKTAIVLLSG
 51 GDSATVAAL AKREGYVIA LSENYGQRH RELRAADIV QALGIDPHFS
 101 INDLAOWMG SSLTDROCTL PQTGVEPDII PSTYVGRNT VFIALGISLA
 151 EAKAEAVFL GINAIDYSG PCRPEYIAT YQOLAALSX VYVGRPIOL
 201 LAPLEISKV DIYHKALEIG VPINOTWCSY AGGEPCGRC DSCRLEQAL
 251 IEAGPELAS AKGRIMREKV D

!!AA:SEQUENCE 1.0
 P1:D69014 - conserved hypothetical protein MTH1108 - *Methanobacterium thermoautotrophicum* (strain Delta H)
 C:Species: *Methanobacterium thermoautotrophicum*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
 C:Accession: D69014
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lum, W.; Pechter, B.; Olu, D.; Spadefora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani, N.; Caruso, G.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrowski, S.; Church, G.M.; Daniels, C.O.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: D69014
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-222 <MTH>
 A:Cross-references: GB:AE000881; GB:AE000665; NID:g2622206; PIDN:AA85597.1; PID:g2622210
 A:Experimental source: strain Delta H
 C:Genetics: GTG
 A:Start codon: conserved hypothetical protein H1191
 C:Superfamily: conserved hypothetical protein H1191
 D69014 Length: 222 February 11, 2000 15:51 Type: P Check: 6260 ..

1 MRAIISIGG MDSAVATAM MDEYEHAIH FDIYGRSRM ELRYARLSE
 51 HLGIEHTLD LOMLGLGGS VLRAGGIPS PSNDDYEC LEFARKWVW
 101 GRNIVFTSIG VSPAEAMDAG AVIYGMDEL AEFIFDNEE FLDAFNLEL
 151 IGLDGVAVY APYIGTKRE IVEAGHEVGL PELIYSCTA GDRHGCYCE
 201 SCMRRRRAFE LAGIDDPTEY RE

!!AA:SEQUENCE 1.0
 P1:C65012 - hypothetical protein b2382 - *Escherichia coli* (strain K-12)
 C:Species: *Escherichia coli*
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
 C:Accession: C65012
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: C65012
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-285 <BLAT>
 A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AMC75441.1; PID:g1788725; UMG:G2382
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: hypothetical protein b2382
 C65012 Length: 285 February 11, 2000 15:51 Type: P Check: 4986 ..

1 MKAPGLPADQ OFEADLFSG VLNPPOLLGRV WFASQASLP VQSLCIDPR
 51 LDIVLGEYG NLEAKOQL VEGEMLEIPA RAANLPVNNK PVMLSLVFA
 101 PTWLGSEFYD SRTSLHPA ROIQLPSLQR GEGEMALTA TLHSRPLEQ
 151 NIIDPLVLSL LHLGRVYVM PGNISOPRGD FLYHSICNV QDNYAOPLTR
 201 ESNAGFINIT PNHSLKLEPQ HGTMREFEY RWRNKAARM ILQYHLISIH
 251 EVAGRCGFDP SDYFCVRFR QFGLTGEGYS ARFQG

!!AA:SEQUENCE 1.0
 P1:A70606 - probable electron transfer protein RV3571 - *Mycobacterium tuberculosis* (strain H37Rv)
 C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 29-Sep-1999
 C:Accession: A70606
 R:Cole, S.T.; Brosch, R.; Parhail, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigelnier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Nature 393, 537-544, 1998
 A:Authors: Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70606
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-358 <CO>
 A:Cross-references: GB:Z922774; GB:AL123456; NID:g3261729; PIDN:CA807145.1; PID:g3066111; PID:g1877302
 A:Experimental source: strain H37Rv
 C:Genetics: K3571
 A:Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferredoxin [2Fe-2S] homology; metalloprotein
 C:Keywords: 2Fe-2S; cytochrome-b5 reductase homology <CBR>
 F:280-344/Domain: ferredoxin [2Fe-2S] homology <FBR>
 F:305-310/313,343/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted
 A70606 Length: 358 February 11, 2000 15:51 Type: P Check: 144 ..

1 MTEALGDEPL GDVLEIOIA EVVDTERDEAR SLVRAVDDGS DDEIPPRRL

```

51  RRAPOGLFLTL RYPERRTGVS ARCYSLCSPS YTDALAVTV KRTADGYSN
101  WLCDAHQVGM RIHVLAPSGN FVPTTLADAF LLLAGSGIT PIMSICKSAL
151  AEGGQAVTLL YANRDRSVI FGDALRELA KYPDLRYLH WLSLQGLPS
201  ASALAKLVAP YTDPRVFCG PGPFQOARD ALAALKVPAQ QVAHEVFKSL
251  ESDPRAAVKY DSGDEAPAT AVELDGOYH TVSPRTAKL LDVLLAGLD
301  APFSCREHC GACACTLRAG KVMGVNDVL EQDLDLGLI LAQSPESD
351  SEVETVDE

!!AA_SEQUENCE 1.0
P1:C65112 - sigma cross-reacting protein 27A - Escherichia coli
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: C65112; JN0287
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
J.: Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: C65112
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-220 <L>A>
A:Cross-references: GB:AE000400; GB:U00096; NID:92367203; PIDN:AAC76241.1;
PID:01789600; WMG:103209
A:Experimental source: Strain K-12, substrain MG1655
R:Ishihama, R.; Fujita, N.; Ishihama, A.
Biochem. Biophys. Res. Commun. 184, 634-639, 1992
A:Title: Identification of Escherichia coli proteins cross-reacting with
antibodies against region 2.2 peptide of RNA polymerase sigma subunit.
A:Reference number: JN0286; MUID:92246944
A:Accession: JN0287
A:Molecule type: Protein
A:Residues: 4-13, 'X', 15-26, 'I' <UES>
A:Comment: This protein reacts with antiserum against a synthetic peptide of
RNA polymerase sigma chains.
C:Genetics:
A:Gene: ybhI
C:Superfamily: sigma cross-reacting protein 27A

C65112 Length: 220 February 11, 2000 15:51 Type: P Check: 597 ..

1  MITMKRIGV LSGGYVDGS EHEAVLILL AISRGAQV CFAPDKQVD
51  VINHLTGEM TETRNVLIEA ARITREIRP LAQDAEELD ALIVPGFGA
101  AKNSNFASL GSECTVDREL KALAQMHA GKPGLFMCIA PAMLPRIEFD
151  PLRLITGDI DTAEVLEMG AEHVPCPVD IVDENKIV TTPAYMLAQN
201  IAEAAGIDK LVSRVYLAE

!!AA_SEQUENCE 1.0
P1:JC5603 - es1 protein - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C>Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 29-Sep-1999
C:Accession: JC5603; PC4488
R:Chang, H.; Gilbert, W.
Biochem. Biophys. Res. Commun. 237, 84-89, 1997
A:Title: A novel zebrafish gene expressed specifically in the photoreceptor
cells of the retina.
A:Reference number: JC5603; MUID:97410119
A:Accession: JC5603
A:Molecule type: mRNA
A:Residues: 1-270 <CHA>
A:Cross-references: GB:U10403; NID:9499129; PIDN:AA60261.1; PID:9499130
A:Experimental source: eye

```

```

A:Accession: PC4488
A:Molecule type: protein
A:Residues: 105-137 <CH2>
C:Comment: This protein is important for maintaining normal retina structure
and function.
C:Genetics:
A:Gene: es1
C:Superfamily: sigma cross-reacting protein 27A

JC5603 Length: 270 February 11, 2000 15:51 Type: P Check: 1627 ..

1  MLASRALIAK QAAAMLVROP ACLMHGGDM GNMGNINIAV VSGCGWMDG
51  TDIHEAAVTM YHLSRNGARF QIFAPNQQM HYMDHAKMP SSDNRNITM
101  ESARFSGQG MQMNDLSKL DANSEDAVIF PGCHGIYKKN STFSKDGKDC
151  KLNVDVERVL KDFHARKPI GLSSMAPLA CRVLPSEVT MGERDESSR
201  WGRPNTNMY QAVKSGNARH NTRPEYAYV DEKNVISTP TFMWETDVHY
251  HYFDGICGM VKHVMKTK

!!AA_SEQUENCE 1.0
P1:T09924 - cytidine deaminase (EC 3.5.4.5) CDA6 - Arabidopsis thaliana
N:Alternate names: Protein T16L4.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Sep-1999
C:Accession: T09924
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.;
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16897
A:Accession: T09924
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <BEV>
A:Cross-references: EMBL:AL079344; GSPDB:GN00062; ATSP:T16L4.120
A:Experimental source: cultivar Columbia; BAC clone T16L4
C:Genetics:
A:Gene: ATSP:T16L4.120
A:Map position: 4
C:Superfamily: cdd protein
C:Keywords: hydrolase

T09924 Length: 293 February 11, 2000 15:51 Type: P Check: 1428 ..

1  MKFVTPSEA AEGVGPSD LPKLDKAMS LARAPVTFK VQAVGLTSSG
51  EYFLGVNVEF PNLPLHHTH AEQFLVTMLA LNSKKLITHI AVSVGTIFG
101  APCGHCQGFY QEMRNAPETE ILIKRPKDI DERMSLSLM PRFPDSIL
151  PEDASLLIQ RDSNLYLSDP EICSDPEDC SHTKRALAA ANKSYAPYSK
201  CPSEVALICG GEYKQWYTE SVAYNPSLGP VEALYDEVA RCGKREFNEI
251  TEYVLVEKMD VAYSGEATAR TFDKIAKPC DFKVLACYRT NKN

!!AA_SEQUENCE 1.0
P1:S5551 - cytokine-inducible protein CIS - mouse
C:Species: Mus musculus (house mouse)
C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 28-May-1999
C:Accession: S5551
R:Toshimura, A.; Ohnubo, T.; Kiyuchi, T.; Jenkins, N.A.; Gilbert, D.J.;
Copeland, N.G.; Hara, T.; Miyajima, A.
EMBO J. 14, 2816-2826, 1995
A:Title: A novel cytokine-inducible gene CIS encodes an SH2-containing protein
that binds to tyrosine-phosphorylated interleukin 3 and erythropoietin
receptors.
A:Reference number: S5551; MUID:95317300
A:Accession: S5551
A:Status: preliminary

```

A:Molecule type: mRNA
 A:Residues: 1-257 <YOS>
 A:Cross-references: GB:D31943; NID:91041128; PIDN:BA06713.1; PID:dl007285;
 PID:q1041129
 C:Superfamily: cytokine-inducible protein CIS; SH2 homology
 F:82-178/Domain: SH2 homology <SH2>
 S55551 Length: 257 February 11, 2000 15:51 Type: P Check: 2728 ..

1 MYLCVQSGCP LLAVEQIGRR PLMQSLLELP GPAMQPLPTG AFPEVTEET
 51 PVQAKNEPKV LDPEGDLCI AKTSYLRES GWWSGITSAS EARGHLOKMP
 101 EGTLYVDST HPSYLFITLSV KTRGPNVR IEVADSFFL DSNCLSRRI
 151 LAFDVAISLV QHYVASCAD TRSDSPDAP TPALPMSKOD APDSVLPID
 201 VATAVHLKLV QPVRSSNR SLOHLCRLVI NRIYAVDCL PLPRRMDYL
 251 RQYPPOL

!!AA_SEQUENCE 1.0
 P1:JC5626 - STAT induced STAT inhibitor 2 - human
 C:Species: Homo sapiens (man)
 C>Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 07-May-1999
 C:Accession: J05626
 R:Matsumoto, S.; Ikegami, K.; Ueno, K.; Narazaki, M.; Naka, T.; Yamamoto, H.;
 Matsumoto, T.; Saito, H.; Hosoe, S.; Kishimoto, T.
 Biochem. Biophys. Res. Commun. 237, 79-83, 1997
 A:Title: Cloning and functional analysis of new members of STAT induced STAT
 inhibitor (SSI) family: SSI-2 and SSI-3
 A:Reference number: J05626; M0ID:97410118
 A:Accession: J05626
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-198 <MIN>
 C:Comment: This protein plays a role in negative feedback control of Janus
 kinase-signal transduction and activator of transcription signaling pathway.
 C:Superfamily: cytokine-inducible protein CIS; SH2 homology
 F:48-149/Domain: SH2 homology <SH2>
 F:163-172/Region: SC motif 1
 F:182-192/Region: SC motif 2
 JC5626 Length: 198 February 11, 2000 15:51 Type: P Check: 1738 ..

1 MTLRCLPESG NGEGTRSQW GTAGSAEPPS PQARLAKAL RELGQTGMW
 51 GSWTVNEAKE KLEAPDGTF LIRDSSHSDY LLTISVYKTA GPTNRIEYQ
 101 DGRFLDSII CVKSKLKQFD SVVHLIDYV QMKCKRTDP EAPRNGTYHL
 151 YLTKPLYTSA PSLOHLCRLT INKCTGAIMG LPLPRLKDY LEEYKFOV

!!AA_SEQUENCE 1.0
 P1:JC5760 - cytokine-inducible SH2 protein 2 - human
 C:Species: Homo sapiens (man)
 C>Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 07-May-1999
 C:Accession: J05760
 R:Matsumoto, H.; Sakamoto, H.; Matsumoto, A.; Suzuki, R.; Yasukawa, H.; Mitsui,
 K.; Wakioka, T.; Tanimura, S.; Sasaki, A.; Mitsuwa, H.; Yokouchi, M.; Ohtsubo,
 M.; Yoshimura, A.
 Biochem. Biophys. Res. Commun. 239, 439-446, 1997
 A:Title: Cloning and characterization of novel CIS family genes.
 A:Reference number: J05760; M0ID:98008857
 A:Accession: J05760
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-198 <MAS>
 A:Cross-references: DDBJ:AB006966
 C:Comment: This protein plays a role in the negative regulation of cytokine
 signaling by interacting with specific targets.
 C:Superfamily: cytokine-inducible protein CIS; SH2 homology
 F:47-142/Domain: SH2 status predicted <SH2>

J05760 Length: 198 February 11, 2000 15:51 Type: P Check: 1381 ..
 1 MTLRCLPESG NGEGTRSQW GTAGSAEPPS PQARLAKAL RELGQTGMW
 51 GSWTVNEAKE KLEAPDGTF LIRDSSHSDY LLTISVYKTA GPTNRIEYQ
 101 DGRFLDSII CVKSKLKQFD SVVHLIDYV QMKCKRTDP EAPRNGTYHL
 151 YLTKPLYTSA PSLOHLCRLT INKCTGAIMG LPLPRLKDY LEEYKFOV

!!AA_SEQUENCE 1.0
 P1:A41900 - cyn operon regulatory protein cynR - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 29-Sep-1999
 C:Accession: A41900; B64761
 R: Sung, Y.C.; Fuchs, J.A.
 J. Bacteriol. 174, 3645-3650, 1992
 A:Title: The Escherichia coli K-12 cyn operon is positively regulated by a
 member of the IySR family.
 A:Reference number: A41900; M0ID:92276346
 A:Contents: K-12
 A:Accession: A41900
 A:Molecule type: DNA
 A:Residues: 1-799 <SUN>
 A:Cross-references: GB:M93053; NID:9145645; PIDN:AA23628.1; PID:9145646
 A:Note: Sequence extracted from NCBI backbone (NCBIN:104623; NCBI:P:104625)
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; M0ID:97426517
 A:Accession: B64761
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-299 <BLAT>
 A:Cross-references: GB:AB000141; GB:U00096; NID:91786532; PIDN:AC73441.1;
 PID:91786533; UMG:P:D0338
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: cynR
 C:Superfamily: probable transcription regulator IySR
 C:Keywords: DNA binding; transcription regulation
 A41900 Length: 299 February 11, 2000 15:51 Type: P Check: 4304 ..

1 MLSRHINFL AVAHEGSFTR AASALHVSOP ALSQOIROLE ESIGVPLFDR
 51 SGRITLITDA GEVWROYASR ALQELGAGKR AIHDVADLRL GSLRIAVPT
 101 FTSYFTIGPLM ADFAARYPSI TLQLOMSQE KIEMLCRDE LDVGIAPAVY
 151 HSPLEFAIPL LTESIALYVA QHHPPLAVHCO VALSRINDER LVLSAEFAT
 201 REQIDHYEER AGHPOVYIE ANSISAVLEL IRRSISITL PAIAIHOHG
 251 LKAISLAPPL LERRAVILRR KNSWQTAAPK AFLMALDKC AYGAGNESR

!!AA_SEQUENCE 1.0
 P1:F70721 - hypothetical protein RV1829 - Mycobacterium tuberculosis (strain
 H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-Aug-1999
 C:Accession: F70721
 R: Cole, S.T.; Brosemer, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
 Gordon, S.V.; Eigmeier, K.; Gars, S.; Barry III, C.E.; Tekle, F.; Badcock, K.;
 Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
 Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.;
 Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
 M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
 Squares, S.
 Nature 393, 537-544, 1998

```

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A/Reference number: A70500; MUID:98295987
A/Accession: F70721
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-164 <COL>
A/Cross-references: GB:278020; GB:AL123456; NID:g3261625; PID:e1300004;
PID:g3261627
A/Experimental source: strain H37Rv
C/Genetics:
A/Gene: RV1829
C/Superfamily: Synecocystis conserved hypothetical protein s111142
F70721 Length: 164 February 11, 2000 15:51 Type: P Check: 7074 ..

1 MGEVAVGIR VEQPQNPVL LREANGDRY LPWIGSEA AAIALQGGV
51 EPPRPETHDL IRDLIALGH SLKEVRIYDL QEGTFVADLI FDRNRYVSAR
101 PSDSVALALR VGVPYIYEEA VLAQAGLLIP DESDEATTA VREDEVEKER
151 EPLDSVSPDD FKAT

!!AA_SEQUENCE 1.0
P1:S56618 - yjx protein - Escherichia coli
C/Species: Escherichia coli
C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 18-Sep-1998
A/Accession: S56618; A65255
R/Burkard, V.; Plunkett III, G.; Sofia, H.D.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A>Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region
from 92.8 through 100 minutes.
A/Reference number: S56314; MUID:95334362
A/Accession: S56618
A>Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-173 <BUR>
A/Cross-references: EMBL:U14003; NID:g1263172; PID:g537234
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August
1994
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burkard, V.;
Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617
A/Accession: A65255
A>Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-173 <BLAT>
A/Cross-references: GB:AE000509; GB:U00096; NID:g2367383; PID:g1790855;
UMGP:04394
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: yjx
C/Superfamily: Escherichia coli conserved yjx protein
S56618 Length: 173 February 11, 2000 15:51 Type: P Check: 3873 ..

1 MLINQVCA TTNPAKIOAI LQAFHEIFGE GSCHIASAVV ESGVDEPFG
51 SEETBAGARN RVANARRLLP EADFVAIEA GIDGSTSTSW VVIENAGRG
101 EARSATLPLP AVILEKVBEG EALGPVMSRY TGIDIGRKE GATGVFTAGK
151 LTRASVYHQA VILALSPFN AVY

!!AA_SEQUENCE 1.0
P1:S45255 - yjx protein homolog - Enterobacter aerogenes (fragment)
C/Species: Enterobacter aerogenes

```

```

C/Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 18-Sep-1998
C/Accession: S45255
R/Arvidson, D.N.; Arvidson, C.G.; Lawson, C.L.; Miner, J.; Adams, C.;
Touderlan, P.
Nucleic Acids Res. 22, 1821-1829, 1994
A>Title: The tryptophan repressor sequence is highly conserved among the
Enterobacteriaceae.
A/Reference number: S45254; MUID:94268903
A/Accession: S45255
A>Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-54 <ARV>
A/Cross-references: EMBL:L26582; NID:g433053; PID:g433055
A/Experimental source: ATCC 13048
A/Note: the nucleotide sequence was submitted to the EMBL Data Library,
December 1993
C/Superfamily: Escherichia coli conserved yjx protein
S45255 Length: 54 February 11, 2000 15:51 Type: P Check: 4167 ..

1 EALGPVMSQH TGIDIGRKE GATGVFTAGK LTRSVYHQA VILALSPFN
51 AIYR

!!AA_SEQUENCE 1.0
P1:A70738 - probable rimI protein - Mycobacterium tuberculosis (strain H37Rv)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Sep-1999
A/Accession: A70738
R/Cole, S.T.; Brosch, R.; Parthill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Baaham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, T.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Horsley, T.; Jorgensen, K.;
Krogan, A.; McLean, J.; Molle, S.; Murphy, L.J.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Nature 393, 537-544, 1998
A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A/Reference number: A70500; MUID:98295987
A/Accession: A70738
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-158 <COL>
A/Cross-references: GB:Z77165; GB:AL123456; NID:g3261609; PID:e255179;
PID:g1449367
A/Experimental source: strain H37Rv
C/Genetics:
A/Gene: rimI
C/Superfamily: Escherichia coli ribosomal-protein-alanine N-acetyltransferase
rimI
A70738 Length: 158 February 11, 2000 15:51 Type: P Check: 8373 ..

1 MPMDEPVTI GALIRADAQR CAELERQLEY GDDPMPAPAF NRELASPHNH
51 YVABSGGTL VGYAGISRLG RPPFEYEVH TIGVPAYOG RGIGRRLLRE
101 LLDFAAGVY YLEVRTDND AALALYSVGF QVGLRRRY RVSGADAYTM
151 RRDGDPDS

!!AA_SEQUENCE 1.0
P1:F72316 - hypothetical protein TM0917 - Thermotoga maritima (strain MS8)
C/Species: Thermotoga maritima
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
A/Accession: F72316
R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft,
D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.;
Ullrich, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.;
Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.;
Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.;

```

Venter, J.C.; Fraser, C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of *Thermotoga maritima*.
 A:Reference number: A72200; MUID:99287316
 A:Accession: F72316
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <ARN>
 A:Cross-references: GB:AE001756; GB:AE000512; NID:94981453; PID:94981454;
 TIGR:TM0917
 A:Experimental source: strain MSB8
 A:Genetics:
 C:Gene: TM0917
 F72316 Length: 309 February 11, 2000 15:51 Type: P Check: 5138 ..

1 MEVLIPSLF LGWSLGANDA ANVGPFGVS GLIPYKRII VASIFVLGS
 51 VLGARGLQN ISSLSISDLI LSIIVLSGA LVTIMTKLG IPVTSQAVV
 101 GGIIQANTV MGIGIDESA LTKILTVFL TPVGAPFLS IFYPLVSLF
 151 RKIPSIQID RVIKISAMII GAYGAFSLGA NNVAANTGVF AKKILITGA
 201 ARLGISIAI GILYKRNW LTVGNLIEL DFTSLIAVL SCAMVWIFS
 251 LIGIHSSSO AIVGAVLAG YKSGMNLGN KVLKILSGV FLTPAVSGTL
 301 SFLTSLIK

11AA_SEQUENCE 1.0
 P1:B70451 - prephenate dehydrogenase - *Aquifex aeolicus*
 C:Species: *Aquifex aeolicus*
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
 C:Accession: B70451
 R:Beckert, G.; Warren, P.V.; Gaasterland, T.; Young, M.G.; Lenox, A.L.; Graham,
 R.D.; Overbeek, R.; Sneed, M.A.; Keller, M.; Anjey, M.; Huber, R.; Feldman,
 R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.
 Nature 392, 553-558, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex*
aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: B70451
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-311 <AOE>
 A:Cross-references: GB:AE000754; NID:92984047; PID:92984048; GB:AE000657
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: tyra
 B70451 Length: 311 February 11, 2000 15:51 Type: P Check: 4885 ..

1 MALISSMNP SPPOGCRKN IIKLKSLSM QNVLIYGVF MGGSFAKSLR
 51 RSGFGRIYV YDINPESISK AYDLGIIDEG TSIKAVEDE SPDEFMLSSP
 101 VRTREIRAKK LSYILSEDAI VTDOGSVKG LVDLENTLG KREYVGHPIA
 151 GTERSGEYS LDIYEGKRV ILPTPKTKDK KRLKIVRWY EDVGVVEYM
 201 SPELDHYVG VVSHLPYAVA FALVDLIHM SPEVDLFKY PGGKDFTR
 251 IAKSDPIWMR DIFLENKRV MKAIEGEFS LNHKLILVR EAEELVEYL
 301 KEVIRKMEI D

11AA_SEQUENCE 1.0
 P1:G39845 - dihydroacetate dehydrogenase (electron transfer subunit) *pyrDII* -
Bacillus subtilis
 C:Species: *Bacillus subtilis*
 C:Date: 20-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 24-Sep-1998

C:Accession: G39845; D69686
 R:Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.
 J Biol Chem 265, 9113-9127, 1991
 A:Title: Functional organization and nucleotide sequence of the *Bacillus*
subtilis pyrimidine biosynthetic operon.
 A:Reference number: A39845; MUID:9125016
 A:Accession: G39845
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <OUT>
 A:Cross-references: GB:M59757; NID:9387576; PID:9143391
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo,
 J.; Bertero, M.G.; Beselieres, F.; Bilotto, A.; Borchert, S.; Boriss, R.;
 Boursier, L.; Brans, A.; Braun, M.; Brignelli, S.C.; Broo, S.; Brouillet, S.;
 Bursch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani,
 J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devigne, K.M.;
 Dueterhoeft, A.; Ehrlich, S.D.; Emmertson, P.T.; Entian, K.D.; Errington, J.;
 Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallazzi,
 A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi,
 G.; Giuseppe, G.; Guy, B.; Haga, K.; Haeche, J.; Hatwood, G.R.; Heaut, A.;
 Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;
 Joris, B.; Karamata, D.; Kasahara, Y.; Kiebert-Biancard, M.; Klein, C.; Kurita, K.;
 Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.;
 Lapidus, A.; Lardinols, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda,
 S.; Mausel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.;
 Nekaj, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega,
 B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
 A.M.; Prescano, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds,
 S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato,
 T.; Scanlon, P.
 A:Authors: Schleich, S.; Schroter, R.; Scofield, F.; Sekiguchi, J.; Sekowska,
 A.; Serró, S.J.; Serró, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccini, E.;
 Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka,
 T.; Terpe, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenoel, M.;
 Vanner, F.; Vassartelli, A.; Viani, A.; Wamburt, R.; Wedler, E.; Wedler, H.;
 Wetzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
 K.; Yata, K.; Yoshida, K.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Authors: Yoshikawa, H.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus*
subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: D69686
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-256 <KUN>
 A:Cross-references: GB:Z99112; GB:AL009126; NID:92633902; PID:9185145;
 PID:92633926
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: pyrDII
 G39845 Length: 256 February 11, 2000 15:51 Type: P Check: 2638 ..

1 MKKAYLVCS NOCIADRVFO MVKLGELYOG FTTPOGFLHL KVEAVTPLL
 51 RRPISADVN FEKNVITILY RVDEGGRILL SIKOGEIYD VLPLNGGFP
 101 VNEVQPKTA LLYGGVGVV PLOELSKRLI EKVAVIIVHL GFOSANDVFX
 151 EEBORQGDY YVATADSYG ETGFTDVIR RKLLEFDILL SCGTPMLKA
 201 LKQYAKREV YLSMERMGCG GIGACFACVC HNESETSYV KVCIDGVPFK
 251 AQEVAL

11AA_SEQUENCE 1.0
 P1:J01382 - hypothetical 34k protein - pea enation mosaic virus
 C:Species: pea enation mosaic virus; PEKV
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 26-Aug-1999
 C:Accession: J01382

R:Demler, S.A.; de Zoeten, G.A.
 J. Gen. Virol. 72, 1819-1834, 1991
 A:Title: The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus.
 A:Reference number: J01382; MUID:91341468
 A:Accession: J01382
 A:Molecule type: genomic RNA
 A:Residues: 1-303 <DEM>
 A:Cross-references: GB:L04573; NID:9294105; PID:9294108
 A:Experimental source: strain MSG
 A:Note: 254-His was also found
 C:Genetics:
 A:Map position: segment RNA1
 C:Superfamily: pea enation mosaic virus hypothetical 34k protein
 J01382 Length: 303 February 11, 2000 15:51 Type: P Check: 6785

1 MHGIGQPOLP LDYVHRCASL SFLLASLDGL LSEARELSGP LALITSSYYL
 51 LVSLALCKAI PGSTWYRPGC WLQPVSGRWL IFCGPTSLAQ RFRLYAARLG
 101 LVLSNCRPH GQSNAILLOS YNALPNNIWM DMAOLDILTF SMDIANTFAY
 151 LADCARPP IVEGVSAYV VPTLGLTHQ DPRLYALRR RNLDLGEPR
 201 RVRQVLESM ALLCSSVRSST SRSROIPLY GSVLHVYGL AERDCILFPT
 251 DSNYSSTYHR VLEODRNRAD QSLFSIDLEY VHDELIALG YSDEDEDID
 301 NFF

11AA_SEQUENCE 1.0
 P1:S56961 - Probable membrane protein YUL178c - Yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein J0490
 C:Species: Saccharomyces cerevisiae
 C>Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 26-Aug-1999
 C:Accession: S56961
 R:Obermaier, B.; Piravandi, E.; Rinke, M.; Dondely, H.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S56937
 A:Accession: S56961
 A:Molecule type: DNA
 A:Residues: 1-196 <OBE>
 A:Cross-references: EMBL:Z49453; NID:91008377; PID:91008378; MIPS:YUL178c
 C:Genetics:
 A:Map position: 10L
 C:Superfamily: Saccharomyces probable membrane protein YUL178c
 C:Keywords: transmembrane protein
 S56961 Length: 196 February 11, 2000 15:51 Type: P Check: 8234

1 MCGGLDVIIL PGKDAITTOI IDPDKNIGFN VEETESALTL TLKGATWGAN
 51 SFAKLEFQC NDNMKODELT SHTWADRSIQ LTLKSGSGL KSKDDKRG
 101 DQNGKDGDS ECKPAPKAG GTSNFWLFL YALLFLIYL MYVEFLNTRG
 151 GSTQDRAEF IQRSTQFLTS LPEFCREYVS RILGRTAQR GGYSAV

11AA_SEQUENCE 1.0
 P1:J01355 - V0 protein - Miscanthus streak virus
 C:Species: Miscanthus streak virus
 C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
 C:Accession: J01355; J00918
 R:Chatant, M.; Matsumoto, Y.; Mizuta, H.; Ikegami, M.; Boulton, M.I.; Davies, J.W.
 J. Gen. Virol. 72, 2325-2331, 1991
 A:Title: The nucleotide sequence and genome structure of the geminivirus miscanthus streak virus.
 A:Reference number: J01355; MUID:92013947
 A:Accession: J01355
 A:Molecule type: DNA
 A:Residues: 1-134 <CHA>

A:Cross-references: DDBJ:D01030; NID:9222128; PID:01001300; PID:9222129
 C:Comment: Miscanthus streak virus causes leaf streak.
 C:Superfamily: Miscanthus streak virus V0 protein
 J01355 Length: 134 February 11, 2000 15:51 Type: P Check: 9413

1 MCGVIMGR GSENNITAP PMPISRPDA SVRCAPGRS LPPLADRDQ
 51 AAVKWRHLS VFALVPLVCG FKRRRLRDR TKGTQVAC LRTWVKFCF
 101 VCKFLIPMV AVAILMNVH CMVYLWHLR CFAS

11AA_SEQUENCE 1.0
 P1:J00150 - hypothetical 13k protein - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Sep-1999
 C:Accession: J00150
 R:Kato, J.; Chu, L.; Kitano, K.; Devault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra, T.K.
 Gene 84, 31-38, 1989
 A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in Pseudomonas aeruginosa characterization of the algR2 gene.
 A:Reference number: J00132; MUID:90108714
 A:Accession: J00150
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-122 <KAT>
 C:Superfamily: Pseudomonas aeruginosa hypothetical 13k protein
 J00150 Length: 122 February 11, 2000 15:51 Type: P Check: 6657

1 MALSPACSR CCAASARP ATACCPFGA SRTCARSRIP WIAMWIRCS
 51 TATVACARS RMPPPKRRT TLPWRACIP NMTTPVTTP MHGRASCWPD
 101 SASASRWSA GSVISPAVCG CA

11AA_SEQUENCE 1.0
 P1:B71217 - hypothetical protein PH2002 - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Sep-1999
 C:Accession: B71217
 R:Kawababayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaebacterium, Pyrococcus horikoshii OT3.
 A:Reference number: A71000; MUID:96344137
 A:Accession: B71217
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-108 <KAN>
 A:Cross-references: GB:AP000007; GB:AP000001; NID:93236134; NID:93236128; PID:01032072; PID:01030009; PID:93258446; PID:93263983
 A:Experimental source: Strain OT3
 A:Note: This accession replaces an interim accession for a sequence replaced by Genbank
 A:Note: This sequence is split into two separate translations in Genbank (residues 1-37, 38-108)
 C:Genetics:
 A:Gene: PH2002
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH2002
 B71217 Length: 108 February 11, 2000 15:51 Type: P Check: 8463

1 MGKHYLEKS ILRPLKIDIF KVCHDPLC EYTLGSGFS LHRPDLFP
 51 VKGSRHPSL VFILKIQGV KEDIFKLLT PSRSQAVRH KPPHOPPEVO

101 SDDPSICP

11AA:SEQUENCE 1.0
 P1:T05820 - hypothetical protein TSK18.150 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
 C:Accession: T05820
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voel, M.; Robben, J.;
 Volckaert, G.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Scheller, C.
 submitted to the Protein Sequence Database, April 1998
 A:Reference number: 215453
 A:Accession: T05820
 A:Molecule type: DNA
 A:Residues: 1-217 <BEV>
 A:Cross-references: EMBL:AL022560
 A:Experimental source: cultured Columbia; BAC clone TSK18
 C:Genetics:
 A:Map position: 4
 A:Insertions: 30/3; 101/2
 A:Note: TSK18.150
 C:Superfamily: Arabidopsis thaliana hypothetical protein TSK18.150

T05820 Length: 217 February 11, 2000 15:51 Type: P Check: 4391

- 1 MHNFLYSV FSGLVSEFT CFAAEFKRTQ KEDIRMDTER NCYVPSGSHAF
- 51 GLGSAAVLCF CLAQIVGNIV VERNHRTRTK REDGYKITDL TLPTVLLLS
- 101 WSNFVVVLLI LSTAFMSRA QAYGEGWIDE DCYLKDGVF AASGLAAILG
- 151 LGALTIASTR IKVKKQOOLV QVYIKDONOD QRSMEEEK HDEHQTNRSE
- 201 SVIHVEVS STNISRI

11AA:SEQUENCE 1.0
 P1:S77462 - hypothetical protein s111071 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Sep-1999
 C:Accession: S77462
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanuma, E.; Nakamura, Y.;
 Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosonoh, S.;
 Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and
 assignment of potential protein-coding regions.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S77462
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-275 <KAN>
 A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018042;
 PID:g1652387
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June
 1996
 C:Superfamily: Synechocystis hypothetical protein s111071

S77462 Length: 275 February 11, 2000 15:51 Type: P Check: 3256

- 1 MANHPSAT RSKNFVWL AASLSLMLW GPASPAQAVN NPELLEKRT
- 51 PVVDLNFEP EIQEALIDD LNSPEVETGW KLRVLTQYDR SPGRVAVPEW
- 101 GLDDKSLIV ADARGNLLA FSIQDEYEL MPTWIMEMQ ARFGNMYYIR
- 151 DNGENLAITE ALETYKGLL KGCNNVPGI PREQWILTVY TSYVGLIFG
- 201 FAALPSPNO TRAWQWYLM SPLMGLIVIA FGIQPVYIRI SDFLEPRNL
- 251 MGFSLGALVA YLSPMFSQIN NNPOI

11AA:SEQUENCE 1.0
 P1:S53572 - probable membrane protein Y11071w-a - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae
 C>Date: 06-May-1995 #sequence_revision 19-Oct-1995 #text_change 20-Sep-1999
 C:Accession: S53572
 R:Smith, V.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: S48407
 A:Accession: S53572
 A:Molecule type: DNA
 A:Residues: 1-158 <SM>
 A:Cross-references: EMBL:Z38060; MIPS:Y11071w-a
 C:Genetics:
 A:Map position: 9L
 C:Superfamily: Saccharomyces probable membrane protein Y11071w-a
 C:Keywords: transmembrane protein
 F:58-74/Domain: transmembrane #status predicted <TM>
 F:102-118/Domain: transmembrane #status predicted <TM2>

S53572 Length: 158 February 11, 2000 15:51 Type: P Check: 4181

- 1 MNKVTLLPP RVFCLSWV WVIDRKSQ RSVNLDQHS NKPPSLSLS
- 51 LLSLSSMSI SLSLLIIV AVLEPPLYS GLPIMSEKS VSFESLSIT
- 101 SFIVPELLI SLTIKIEL SSKCITPSI SKENCPIIN FVNSLSISK
- 151 SIPEVDCK

11AA:SEQUENCE 1.0
 P1:G45355 - ORF1 protein - Autographa californica nuclear polyhedrosis virus
 C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
 C>Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 09-Sep-1997
 C:Accession: G45355
 R:Guarino L.A.; Smith, M.W.
 Virology 179, 1-8, 1990
 A:Title: Nucleotide sequence and characterization of the 39K gene region of
 Autographa californica nuclear polyhedrosis virus.
 A:Reference number: A45355; MUID:91021010
 A:Accession: G45355
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <GUA>
 A:Cross-references: GB:M37122; NID:g332383; PID:g332390

G45355 Length: 146 February 11, 2000 15:51 Type: P Check: 1335

- 1 MNTLOOPLY AYAVHGAAR TKIAFDIDG TLISSTREX FPKNPDDMQL
- 51 LPCAKRLRL YELGYDIYF TNOAHLGSGK IKASDLIYKL ENIKKATGV
- 101 ISFYSPKND EHRKPTREGG AHPNSLRIL TRNNLFMWT RLAEIL

11AA:SEQUENCE 1.0
 P1:S32975 - gene BCRF2 protein - human herpesvirus 4
 N:Alternate names: gene BMRF1 protein
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Mar-1998
 C:Accession: S32975; S32976; S32977; S32978; S32980; S32981; S32982;
 S32983; S32984; S32985; S32986
 R:Farrell P.J.
 submitted to the EMBL Data Library, March 1988
 A:Reference number: S32975
 A:Accession: S32975
 A:Molecule type: DNA
 A:Residues: 1-383 <FAV>
 A:Cross-references: EMBL:V01555; NID:g59074; PID:g1334836;
 PID:g25078; PID:g1334837; PID:g25079; PID:g1334838; PID:g25080; PID:g1334839;
 PID:g25081; PID:g1334840; PID:g25066; PID:g1334841; PID:g25067; PID:g1334842;
 PID:g25068; PID:g1334843; PID:g25069; PID:g1334844; PID:g25070; PID:g1334845;
 PID:g25071; PID:g1334846; PID:g25072; PID:g1334847
 A:Note: each of the twelve author supplied translations in EMBL:V01555 for this

repeated gene is marked as conflicting with the conceptual translation

C:Genetics:
A:Gene: BCR2_1; BMRFL_2; BMRFL_3; BMRFL_4; BMRFL_5; BMRFL_6; BMRFL_7; BMRFL_8;
BMRFL_9; BMRFL_10; BMRFL_11; BMRFL_12
A:Note: twelve consecutive ORFs apparently encode the identical polypeptide
S32975 Length: 383 February 11, 2000 15:51 Type: P Check: 4975

1 VWEAGRRPR GEVEDRRPGL CMQSPGDPR PCGPRSPSA PQDPRPSRQ
51 GRASSGAAGS PPOAPOTRVS ASRADRRAM RLIGASRRGW FCPSLCPSE
101 PGTSGTPEPL GPASRRPGL RSLSPVKPK ECLNGATLGA QAPESGQGH
151 LRVPRVPGQ PEGPRPGRP QRPVPRFPFG LQSPGCPPEG TLGVPSPLO
201 ARASSRRGA SLGVOVPHR DSGDPPTG PSICPPAPLO PSILHPPQL
251 ASPGPPQPE GPRPGRVAF PLPWLPLAS HPSPLSLPH RVHQGRDP
301 GGPVSVPPAA AQLPPGKGA SPSPSLPS LLCTVCKVOP PTPVHGRSQ
351 PRPLPVRDR PSVHPGHRP RVSTVPSPRG DFM

!!AA_SEQUENCE 1.0
P1:S02383 - Probable membrane antigen C13 - human herpesvirus 4 (fragment)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 07-Feb-1997
R:Wallis, D.; Gannon, F.
EMBO J. 7: 1191-1196, 1988
A:Title: The expression of novel antigens from the Epstein-Barr virus large
internal repeat
A:Reference number: S02381; M01D:88296424
A:Accession: S02383
A:Molecule type: DNA
A:Residues: 1-57 <WAL>
A:Cross-references: EMBL:X07816
C:Keywords: membrane protein; surface antigen
S02383 Length: 57 February 11, 2000 15:51 Type: P Check: 5478

1 RGNFCSLCP SEEPCTGTP EPIGPASRRP PGLRSPSPV KPRCLNGAT
51 LGQAPE

!!AA_SEQUENCE 1.0
P1:S03243 - hypothetical protein B-115 - Sulfolobus particle SSV1
C:Species: Sulfolobus particle SSV1
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 18-Jun-1993
R:Palin, P.; Grampp, B.; Zillig, W.; Yeats, S.; McWilliams, P.; Relter, W.D.
submitted to the EMBL Data Library, March 1988
A:Reference number: S03243
A:Accession: S03243
A:Molecule type: DNA
A:Residues: 1-115 <PAL>
A:Cross-references: EMBL:X07234
S03243 Length: 115 February 11, 2000 15:51 Type: P Check: 3042

1 MTEFNANSIR AKILRKILQ LAENVYISA SLISHTLLS VATVLRHRI
51 LNDCEYIELY KQCRITYANI RDNARQIQL NSELEGFKNV SKPILTKDE
101 TPKEFGKDS LNRG

!!AA_SEQUENCE 1.0
P1:S21504 - hypothetical protein (gene 29 5' region) - phage SP01
C:Species: phage SP01
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C:Accession: S21504
R:Wilhelm, K.; Koch, T.; Rueger, W.

submitted to the EMBL Data Library, June 1991
A:Description: deoxyribose hydroxymethylase gene of phage SP01 Bacillus
subtilis.
A:Reference number: S21504
A:Accession: S21504
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-148 <WIL>
A:Cross-references: EMBL:X60728; NID:914869; PID:914870
S21504 Length: 148 February 11, 2000 15:51 Type: P Check: 7782

1 MNOOMVLIK GVVVDPOVK VESKYSLE VTKKVVHRA LDIITPGLA
51 TKLMALSPG ALAAGVDSA DKIRGFHDI IDVFTALAE ILMFYALTAC
101 VLIATKKNKA GWERLKNVGY AVAGIALIPT FSEFLRWSS IVSSITF

!!AA_SEQUENCE 1.0
P1:S35282 - eaa protein - phage P22
C:Species: phage P22
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 07-May-1999
C:Accession: S35282
R:Wulff, D.L.; Ho, Y.S.; Powers, S.; Rosenberg, M.
Mol. Microbiol. 9, 261-271, 1993
A:Title: The int genes of bacteriophages P22 and lambda are regulated by
different mechanisms
A:Reference number: S35280; M01D:94018622
A:Accession: S35282
A:Molecule type: DNA
A:Residues: 1-317 <WUL>
A:Cross-references: EMBL:L06296; NID:9215287; PID:9215290
C:Genetics:
A:Gene: eaa
S35282 Length: 317 February 11, 2000 15:51 Type: P Check: 7758

1 MTTITERIE LTVKSPLENG LTRGEOMEIA RIALASLDAE TVRYLNKFSG
51 TCVTLEOQPN AADVAVYIP LVAAVPPER ERIRREHAW SDRTPDDYCP
101 VGPLKHLNKE ALFAADPSD PLENDMCP LMDQRNNGI SDFITRANI
151 EKLEINKTRO WPEKDCPEP LHKROPEV VPECCPAELP YQVKAVALD
201 VALCQSGEV VTYTPDEKA TIWINNSGT CYQEVYLER LQALAGNSP
251 VIPGWNICS ERMNDDESK PLAITGKCL GQGFVATYD DQGFEDYWG
301 MEIIVSHM QLPDPPL

!!AA_SEQUENCE 1.0
P1:S17834 - acetyl-CoA carboxylase - phage T7
C:Species: phage T7
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S17834
R:Lopez-Castillas, F.; Kim, K.H.
Eur. J. Biochem. 201, 119-127, 1991
A:Title: The 5' untranslated regions of acetyl-coenzyme A carboxylase mRNA
provide specific translational control in vitro.
A:Reference number: S17834; M01D:92007865
A:Accession: S17834
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-36 <LOP>
S17834 Length: 36 February 11, 2000 15:51 Type: P Check: 1383

1 MLDQKGTGV QIRPLCLSR ATVFGKLTG LMSDQ
!!AA_SEQUENCE 1.0
P1:A56663 - capsid protein VP1 - foot-and-mouth disease virus Asia (strain Asia
1/63/72)

C:Species: Apthovirus Asia (foot-and-mouth disease virus Asia)
C>Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 29-Jan-1999
C:Accession: A56663
R:Surayanarayana, V.V.; Rao, B.U.; Padayatty, J.D.
Indian J. Biochem. Biophys. 29, 20-24, 1992
A:Title: Nucleotide sequence of the cDNA and the derived amino acid sequence for the major antigenic protein of foot and mouth disease virus, type Asia 1 63/72.
A:Reference number: A56663; MUID:92275633
A:Accession: A56663
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <SOR>
A:Cross-references: GB:S37121; NID:q249692; PID:q249693
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:104727, NCBI:P:104728)
A56663 Length: 299 February 11, 2000 15:51 Type: P Check: 6051 ..

1 MTNCCRGGL DPGKRRHVLQ GRVAVFIAGD ADILRLHDA GHCHRSQCT
51 RYALDLVNRQ FAGRQORAR RSRRCETPA GSARACRARR RSSPRHSGP
101 CHRASAPMSI AVCPVFSMKM ATRVSPSCAA TRPSIAKGP NMGDIAAIL
151 RVADRRLIDE DLEELVIDD ALPIGPDDR DLAGORIGAA HAIDLARVGR
201 SHGSOOEVA RGAVGGQVFR KEIALMGRAA AHPHDEGAGR DVRHISGPP
251 AGAGAYCRR GPALRRPGR ANLPFRPRAC AFSSTAASA RGMHPPPL

11AA_SEQUENCE 1.0
P1:H72245 - 2,3,4,5-tetrahydropyridine-2-carboxylate
N:succinyltransferase-related protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: H72245
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Uitterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
A:Reference number: A72200; MUID:99287316
A:Accession: H72245
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-236 <ARN>
A:Cross-references: GB:AE001799; GB:AE000512; NID:q4982067; PID:q4982085;
A:Reference number: A72200; MUID:99287316
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TMI519
H72245 Length: 236 February 11, 2000 15:51 Type: P Check: 6860 ..

1 MSRLDRETI EMIAKAKKT PIYAIKGL AGIDSSKRF FQDERGILF
51 GEYEDFKLL EHRREKIEDY HLEVARNNSA LPLADLTXX ARLEPGIIR
101 DWVEIGGAV IMGAVINVG AVIGECIMD MNAYVGGRAI IGKCHIGAG
151 AVIAGVIEPP SAKPVIEDE VLVGANAVIL EGVYGGKAV VAAGAVYTKD
201 VPPYTVAGV PARVIKQIDE KTRKTIKIVD ELRNLE

11AA_SEQUENCE 1.0
P1:E72316 - hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

C:Accession: E72316
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Uitterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.
A:Reference number: A72200; MUID:99287316
A:Accession: E72316
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <ARN>
A:Cross-references: GB:AE001757; GB:AE000512; NID:q4981464; PID:q4981479;
A:Reference number: A72200; MUID:99287316
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0941
E72316 Length: 410 February 11, 2000 15:51 Type: P Check: 6581 ..

1 MYRISFLIL ILSSLLFSK EVLENTYNT GIVSLNRTF PFETSRDLF
51 FKASLEFIYT PKVNGITID LPSITLLVLS NGVIGVDVG DIRIPEDVEA
101 ELLESTDLAF YDAFPEVEF SDRVYVKS ITREKLEYLL SFEFKIEVP
151 RRIEGTIDVT PKPEVRSXK VMHPLIGTVY FLNVYDSDTF VILMNGREGK
201 LWTQTEATK LEVEVDSIG QSTSLSEAL PVSFERSYQ SSVEFGOSLI
251 LPLVWVFELP SNGEITVFS PEPFGVYHLI GCDSEGNFLR MEIDVRDTP
301 PVLSPKIE DASSPVEYI CDGREAOKIP DGHVVEVKA TDFGNTSTA
351 FEFVNRHRY IYRERPVY LSPKRIQID GLSLGSLIT GWIREVEVM
401 VGEVYKLEK

11AA_SEQUENCE 1.0
P1:B40609 - probable lipoprotein MIPA - Myxococcus xanthus
C:Species: Myxococcus xanthus
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C:Accession: B40609
R:Martinez-Camero, M.; Munoz-Dorado, J.; Fares-Vidal, E.; Inouye, M.; Inouye, S.
J. Bacteriol. 175, 4756-4763, 1993
A:Title: Car, a 115-kilodalton membrane protein required for development of Myxococcus xanthus.
A:Reference number: A40609; MUID:93328680
A:Accession: B40609
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-236 <MAR>
A:Cross-references: GB:S64103; NID:q402803; PID:q402805
A:Note: sequence extracted from NCBI backbone (NCBIN:135644, NCBI:P:135646)
B40609 Length: 236 February 11, 2000 15:51 Type: P Check: 8809 ..

1 MKNIVNTAL VLVGASLIT GCNFEOPETN CEFQESPWA VKYDVVSPK
51 DANGDECTT APVLEMGVY KYVNPETGAA QALRPATIA SRAIADTTT
101 SADOTISGL DTEPKDHGC HANDEAPAFV NVAASDTAA NTIYEFTNV
151 RYSAVAAPG TQFTGELKTI SNGTSSVYM RAVWPAPCD TASTEPANEC
201 GVSGINPEF AVYCPITSAT GTIGCVFAG DIPSEK

11AA_SEQUENCE 1.0
P1:JC6019 - response-regulator protein Frz2 - Myxococcus xanthus

C:Species: Myxococcus xanthus
 C:Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 07-May-1999
 C:Accession: J66019
 R:Rutland K.G., Ward, M.J., Zusman, D.R.
 Mol. Microbiol. 20, 645-653, 1996
 A:Title: Identification and characterization of FrzZ, a novel response regulator necessary for swarming and fruiting-body formation in Myxococcus xanthus.
 A:Reference number: J66019; MUID:96347136
 A:Accession: J66019
 A:Molecule type: DNA
 A:Residues: 1-290 <TRU>
 C:Cross-references: GB:U47814; NID:q1236916; PID:q1236917
 C:Comment: This protein belongs to the response-regulator superfamily, it is involved in the chemotactic signal-transduction pathway that functions in a manner analogous in enteric bacteria, it regulates the individual cell reversal, responds to attractants and repellents in the spatial chemotaxis assay, and it is necessary for swarming and fruiting-body formation.
 C:Genetics:
 A:Gene: frzZ
 C:Superfamily: response regulator homology
 C:Keywords: phosphoprotein; signal transduction
 F:4-114/Domain: response regulator homology <RRH1>
 F:124-162/Region: alanine-proline-rich
 F:170-283/Domain: response regulator homology <RRH2>
 F:32/Binding site: phosphate (Asp) (covalent) #status predicted
 F:220/Binding site: phosphate (Asp) (covalent) #status predicted
 J66019 Length: 290 February 11, 2000 15:51 Type: P Check: 2747 ..

1 MSRVLYIDS PMLVLYTRA LTAAGYASG AODLASLEIK LAEPFSLIT
 51 MOVNPEMG DDVVEYLKQ KYTAKLVLY SPISSEALDG KTKASGADGY
 101 ILSGGLNAV LGVWGLIGP PALSIPTAV APAAPAAP TPAPAPAPA
 151 PAATGLAKA PTTGGRKPR LIYDSEMTA RIEADLYTK GFEVHVADTA
 201 DKATILKK QTRPDVLLD VRMPVNGRG PCRFKNSL FKGIKVLCS
 251 GENVELORI CREAGADGY PKDVGNGLV AKELMPTGNE

!!AA_SEQUENCE 1.0
 P1:C70198 - conserved hypothetical protein BB0788 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
 C:Accession: C70198
 R:Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, B., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, A.R., Quackenbush, J., Salzberg, S., Hanson, M., Vugt, R.V., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Uitterback, T., Wathey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: C70198
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-440 <TRU>
 A:Cross-references: GB:AE001177; GB:AE000783; NID:92688711; PID:92688713;
 TIGR:BB0788
 A:Experimental source: strain B31
 C70198 Length: 440 February 11, 2000 15:51 Type: P Check: 1619 ..

1 MHFLDENIQI KIDKFKKNS LDKNRVYAF SGGADSTALL INLKYYLSNN
 51 VIAFYEAHT RSDNEDONEI EHVAGFCDLY NIALQIKKD IDIKESARL
 101 GVSIEELAK FYVIALENL KENGANYIAL AHNEDQIEI IMRFQGSF

1 LDGSGIPSV NRIIRPLE VSRLEIENFL SLNNIGFEVD STNAONLYL
 201 NRVANNLLPA IKVFEKGYEK CIKRISEPSK EPADYEGKDE FEPVKGKRY
 251 KSFOLKTFD FKVLYFRLI FKILNSGIA AKVSYALNE AKVEINRKK
 301 NNVLKTNDF ELEKRNKTN LIFKDEKEY KPDFLILEVG KWHSLSGKI
 351 LKYLECNA SVSRLEKCSY EFRYKFKDR LKAKKFEKSF IRCNFAYML
 401 LALNRLIGI IDLNTNLVW SEKSILKRN ISLIGLKE

!!AA_SEQUENCE 1.0
 P1:E70167 - hypothetical protein BB0542 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
 C:Accession: E70167
 R:Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, B., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, A.R., Quackenbush, J., Salzberg, S., Hanson, M., Vugt, R.V., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Uitterback, T., Wathey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: E70167
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-192 <TRU>
 A:Cross-references: GB:AE001156; GB:AE000783; NID:92688459; PID:92688469;
 TIGR:BB0542
 A:Experimental source: strain B31
 E70167 Length: 192 February 11, 2000 15:51 Type: P Check: 2658 ..

1 MDNLNSINSI KDLSEFMYK LLEIIVFVLS CSSEIFKRYON ISGEYKLLAK
 51 LNEELGNET SVLIERKIK FNINAGDSS YNFIACINL KRVVELEKL
 101 NSIIEEDEN ILINKGYL LFRKNDLNA LIYIKLIEF APANKREALFN
 151 IFYIHLKSD KKNKKYISK YKELNHPIS GAEEIVSIL KS

!!AA_SEQUENCE 1.0
 P1:B70174 - hypothetical protein BB0595 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
 C:Accession: B70174
 R:Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R., Lathigra, B., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, A.R., Quackenbush, J., Salzberg, S., Hanson, M., Vugt, R.V., Palmer, N., Adams, M.D., Gocayne, J., Weidman, J., Uitterback, T., Wathey, L., McDonald, L., Artach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: B70174
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-208 <TRU>
 A:Cross-references: GB:AE001161; GB:AE000783; NID:92688515; PID:92688526;
 TIGR:BB0595
 A:Experimental source: strain B31
 B70174 Length: 208 February 11, 2000 15:51 Type: P Check: 8103 ..

```

1 MOKKEFYNTN KYKSHMLMF PIFAATKNFL DASIASWIS ICILPALII
  (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 16-Oct-1998
C:Accession: H64586
R:Gomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Rieschmann, R.D.; Kirschum, K.A.; Klein, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, R.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
B.; Richardson, D.; Dodson, R.; Khalek, H.G.; Glodek, A.; McKenney, K.;
Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Cocayne,
J.D.; Utterback, I.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman,
J.M.; Fujii, C.; Bowman, C.; Wathey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: H64586
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-114 <TOM>
A:Cross-references: GB:AE000568; GB:AE000511; NID:92313646; PID:92313651;
TIGR:HP0536

H64586 Length: 114 February 11, 2000 15:52 Type: P Check: 7347 ..

1 MKRPISKLKQ NLFQKHSFN KHLDKSYLY RLFNISSIVI GLVALFSYG
51 AGVILVYPII FLFALIKPS FFYTYTLL LVSLSISKY YLSHANFTM
101 MNGIGNLMS QMNGDYPNN PFMYGYAPTY SDSSTLPIL GX

11AA_SEQUENCE 1.0
PI:H64587 - cag pathogenicity island protein cag21 - Helicobacter pylori
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 16-Oct-1998
C:Accession: F64587
R:Gomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Rieschmann, R.D.; Kirschum, K.A.; Klein, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, R.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
B.; Richardson, D.; Dodson, R.; Khalek, H.G.; Glodek, A.; McKenney, K.;
Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Cocayne,
J.D.; Utterback, I.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman,
J.M.; Fujii, C.; Bowman, C.; Wathey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: F64587
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-142 <TOM>
A:Cross-references: GB:AE000568; GB:AE000511; NID:92313646; PID:92313651;
TIGR:HP0542

F64587 Length: 142 February 11, 2000 15:52 Type: P Check: 2551 ..

1 MKNFKYKIL LFAMCLIGM FNAPLNADON TDIKDISPED MALNSVGLVS
51 RDQLKEIPK ETLQKXAVL NDYNDKNVNI KFDNISLGSF QPNDNLGINA
101 MNGIGNLMS QMNGDYPNN PFMYGYAPTY SDSSTLPIL GX

11AA_SEQUENCE 1.0
PI:F64587 - cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: F71925
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;
Smith, D.R.; Noonan, B.; Gill, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.;
Caruso, A.; Ulla-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.;
Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557
A:Accession: D71926
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <ARN>
A:Cross-references: GB:AE001482; GB:AE001439; NID:94155018; PID:94155024
A:Experimental source: strain J99
C:Genetics:

D71926 Length: 142 February 11, 2000 15:51 Type: P Check: 2840 ..

1 MKNFKYKIL LFAMCLIGM FNAPLNADON TDIKDISPED MALNSVGLVS
51 RDQLKEIPK ETLQKXAVL NDYNDKNVNI KFDNISLGSF QPNDNLGINA
101 MNGIGNLMS QMNGDYPNN PFMYGYAPTY SDSSTLPIL GX

11AA_SEQUENCE 1.0
PI:F71925 - cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: F71925
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;
Smith, D.R.; Noonan, B.; Gill, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.;
Caruso, A.; Ulla-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.;
Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557
A:Accession: F71925
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-114 <ARN>
A:Cross-references: GB:AE001482; GB:AE001439; NID:94155018; PID:94155033
A:Experimental source: strain J99
C:Genetics:

A:Gene: cagP

F71925 Length: 114 February 11, 2000 15:52 Type: P Check: 6963 ..

1 MKRPISKLKQ NLFQKHSFN KHLDKSYLY RLFNISSIVI GLVALFSYG
51 AGVILVYPII FLFALIKPS FFYTYTLL LVSLSISKY YLSHANFTM
101 KLIIMTQWQ NMFL

11AA_SEQUENCE 1.0
PI:H64586 - cag pathogenicity island protein cag15 - Helicobacter pylori
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 16-Oct-1998
C:Accession: H64586
R:Gomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Rieschmann, R.D.; Kirschum, K.A.; Klein, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, R.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
B.; Richardson, D.; Dodson, R.; Khalek, H.G.; Glodek, A.; McKenney, K.;
Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Cocayne,
J.D.; Utterback, I.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman,
J.M.; Fujii, C.; Bowman, C.; Wathey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: H64586
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-114 <TOM>
A:Cross-references: GB:AE000568; GB:AE000511; NID:92313646; PID:92313651;
TIGR:HP0536

H64586 Length: 114 February 11, 2000 15:52 Type: P Check: 7347 ..

1 MKRPISKLKQ NLFQKHSFN KHLDKSYLY RLFNISSIVI GLVALFSYG
51 AGVILVYPII FLFALIKPS FFYTYTLL LVSLSISKY YLSHANFTM
101 KLIIMTQWQ NMFL

11AA_SEQUENCE 1.0
PI:H64587 - hypothetical protein HP1033 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997
C:Accession: A64549
R:Gomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;

```

```

Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.;
Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne,
J.D.; Uterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman,
J.M.; Fujii, C.; Bowman, C.; Matthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: A64649
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-131 <TOM>
A:Cross-references: GB:AE000611; GB:AE000511; NID:92314173; PID:92314186;
TIGR:HP1033

A64649 Length: 131 February 11, 2000 15:52 Type: P Check: 6457 ..

1 MYSKILATSF SLPLNSWMLR SKLSVPPLS KQISKSVOA IKIDANNNAI
51 KYPSPVDNTV HNRMLSGSIN FNTENTNPKR NPTTEKWK FCTETQAYPK
101 YKSLKRLF KPSFGISKAP VSNLETKRA I

11AA_SEQUENCE 1.0
P1:D64710 - hypothetical protein HP1524 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997
C:Accession: D64710
R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus,
B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.;
Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne,
J.D.; Uterback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman,
J.M.; Fujii, C.; Bowman, C.; Matthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: D64710
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-115 <TOM>
A:Cross-references: GB:AE000650; GB:AE000511; NID:92314700; PID:92314706;
TIGR:HP1524

D64710 Length: 115 February 11, 2000 15:52 Type: P Check: 2117 ..

1 MRSKITHFLA ISFVLSLSA CKDEPKSSQ SHQNTKTKR NNDINQAND
51 IRIIEHEED EKATREYNDL INNENKIDEI NNEENAPDSQ KRTNNVLAQA
101 TNHODNLSP LNRRY

11AA_SEQUENCE 1.0
P1:E71938 - hypothetical protein Jhp0391 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: E71938
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;
Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.;
Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.;
Mills, S.D.; Jiang, O.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557
A:Accession: A71809
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-307 <ARN>
A:Cross-references: GB:AE001491; GB:AE001439; NID:94155127; PID:94155142
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0581

E71938 Length: 307 February 11, 2000 15:52 Type: P Check: 428 ..

1 MKKVLTLTSLISFWLHAE RNFYLGINF ABSYIQOGG SIGKASAE
51 IRIIEHEED EKATREYNDL INNENKIDEI NNEENAPDSQ KRTNNVLAQA
101 TNHODNLSP LNRRY

11AA_SEQUENCE 1.0
P1:C71912 - probable outer membrane protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
C:Accession: C71912
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;
Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.;
Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.;
Mills, S.D.; Jiang, O.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.
A:Reference number: A71800; MUID:99120557
A:Accession: C71912
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-307 <ARN>
A:Cross-references: GB:AE001491; GB:AE001439; NID:94155127; PID:94155142
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0581

C71912 Length: 307 February 11, 2000 15:52 Type: P Check: 428 ..

1 MKKVLTLTSLISFWLHAE RNFYLGINF ABSYIQOGG SIGKASAE
51 IRIIEHEED EKATREYNDL INNENKIDEI NNEENAPDSQ KRTNNVLAQA
101 TNHODNLSP LNRRY

```

```

51 ALNQAINNAR NSLFPNTKAI RDVONALNAV KDSNKIANEF AGNGSGGIE
101 NELSAGKYF LCKKRIGFR HSLFFGYOLG GVGSVPGSL IAPLPYGMT
151 DLINMTNDR RASOEYERR VKGLSIFYKD MGRRLTADT LKRASRIIR
201 KSSGLVIGME LGASTWFASN NLTPFNOVRS RTIFOLQKF GVRFSDEYD
251 IDRGDGNLY GGSVELGVK VPAFKNVYS DDYGRDLXK RVSVYLYNTY
301 YNFKKH

!!AA_SEQUENCE 1.0
P1:PH0856 - mauF protein - Paracoccus denitrificans
C:Species: Paracoccus denitrificans
C>Date: 17-Aug-1992 #sequence_revision 26-Apr-1996 #text_change 07-May-1999
C:Accession: S51048; PH0856
R:van Spanning, R.J.M.; van der Palsen, C.J.N.M.; Slotboom, D.J.; Reijnders,
W.N.M.; Stouthamer, A.H.; Dulle, J.A.
Eur. J. Biochem. 226, 201-210, 1994
A:Title: Expression of the mau genes involved in methylaniline metabolism in
Paracoccus denitrificans is under control of a lysR-type transcriptional
activator.
A:Reference number: S51046; MUID:95045590
A:Accession: S51048
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <VAN>
A:Cross-references: EMBL:012464; NID:9558801; PID:9558804
R:Chistostedov, A.Y.; Boyd, J.; Mathews, F.S.; Lidstrom, M.E.
Biochem. Biophys. Res. Commun. 184, 1181-1189, 1992
A:Title: The genetic organization of the mau gene cluster of the facultative
autotroph Paracoccus denitrificans.
A:Reference number: PH0856; MUID:92272706
A:Accession: PH0856
A:Molecule type: DNA
A:Residues: 129-277 <CHT>
A:Cross-references: GB:M90099; NID:9150583; PID:9150584
C:Genetics:
A:Gene: mauF

PH0856 Length: 277 February 11, 2000 15:52 Type: P Check: 7474 ..

1 MVSVEDLHGL SAGQASVPPC KLFQSPSPSA TRINVLAA LAGAAGVAL
51 ASAGPOPLW AVLGAANVAG GLSTWSPCG YSSISLIRPD GGLNAVAGW
101 LPTFAMHAG YGLGALMLGG LIGIGLIGAG FSGEGSTALL VGLVGLANG
151 AHQIDELRVP YPQRAQVPH DARQRPKWV IGLYGLSLG LDLYTYVOTP
201 LLYMTLAAV FTGNIAHAA IVALFNLGRF LPVAVNALPT PDYVOAMLA
251 RHOENALAD GAILTALGAG FTVLALI

!!AA_SEQUENCE 1.0
P1:19703 - tram protein - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 19-Jul-1996
C:Accession: 19703
R:Funua, C.; Burbee, M.; Minns, S.C.
J. Bacteriol. 177, 1367-1373, 1995
A:Title: Activity of the Agrobacterium T1 plasmid conjugal transfer regulator
Trar is inhibited by the product of the tram gene.
A:Reference number: 19703; MUID:95173115
A:Accession: 19703
A:Status: preliminary; translated from GB/EMBL/DDAB
A:Molecule type: DNA
A:Residues: 1-102 <RES>
A:Cross-references: EMBL:U016786; NID:9571511; PID:9571512
C:Genetics:
A:Gene: tram

139703 Length: 102 February 11, 2000 15:52 Type: P Check: 4597 ..
1 MELEDANVK KVELRPLGL TRGLEPTDLE TTITDAIRH RALVKADEL
51 FOALPEYTKT GOACGPOHI RYEASIEBH AQMSALNTLY SLGFIPRVV
101 VN

!!AA_SEQUENCE 1.0
P1:JC2568 - mirax protein - Rhizobium meliloti (fragment)
C:Species: Rhizobium meliloti
C>Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 14-Jul-1995
C:Accession: JC2568
R:Leach, F.; Wacks, D.B.; Signer, E.R.
Gene 148, 87-90, 1994
A:Title: Rhizobium meliloti homologs of Escherichia coli mur genes.
A:Reference number: JC2567; MUID:95011665
A:Accession: JC2568
A:Molecule type: DNA
A:Residues: 1-118 <LEA>
A:Cross-references: GB:L25875
C:Genetics:
A:Gene: mirax

JC2568 Length: 118 February 11, 2000 15:52 Type: P Check: 7036 ..
1 LQIHVPPTG ELAVITLAVI GGIGFLMFNA PPAIFMGDT GSIALGASIA
51 PLAAATKHEI VMTIGGLEV IETLSVITOV FWFKRTGHRV FLIGADPPSL
101 REEGLDKRPQ GDPFWMQ

!!AA_SEQUENCE 1.0
P1:S27344 - hupK protein - Rhizobium leguminosarum
C:Species: Rhizobium leguminosarum
C>Date: 25-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 17-Mar-1999
C:Accession: S27344
R:Rey, L.; Hidalgo, E.; Palacios, J.; Ruiz-Arguieso, T.
J. Mol. Biol. 228, 998-1002, 1992
A:Title: Nucleotide sequence and organization of an H(2)-uptake gene cluster
from Rhizobium leguminosarum bv. viciae containing a rubredoxin-like gene and
four additional open reading frames.
A:Reference number: S27340; MUID:93108466
A:Accession: S27344
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <REV>
A:Cross-references: EMBL:X52974; NID:91167855; PID:948731

S27344 Length: 370 February 11, 2000 15:52 Type: P Check: 1548 ..
1 MFLLAGITI GIDTVSRAL ACSVAVKANR PRGLTBMVG ROPEEAPVLA
51 GOVFSICGFR QSVARLAVL AADDLAMDE ERLGASAGLI AEFIFETLRA
101 LILQWPTPLP ERFADAGRH LREALASIA IISRAKAGT SRPRLAAAE
151 RLSAATATLG IPGRGDTPLP ETACAAILD VEDDHVAGR RPDPLTISD
201 AEVVARLRDE AGYASLPHLS GRIETGAYA RSASAGLPEA PHLARLRAR
251 IGDVRISLNO LTAARTGDF DCASLACSGP TPGAQGYAV ECARGRLYHQ
301 IEIGDGRILA AYRIAPTEM NFHPAGPVE TLLSSPVGD EAAVRISRL
351 AVLEDPVAF EINVREADA

!!AA_SEQUENCE 1.0
P1:S28677 - hypothetical protein 4 - Rhizobium sp. (strain IC342)
C:Species: Rhizobium sp.
C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S28677

```


R:Upadhyaya, N.M.; Scott, K.F.; Tucker, W.T.; Watson, J.M.; Dart, P.J.
 M01: Plant Microb. Interact. 5: 129-143, 1992
 A:Title: Isolation and characterization of *Rhizobium* (IC3342) genes that
 determine leaf curl induction in pigeon pea.
 A:Reference number: S28673
 A:Accession: S28677
 A:Molecule type: DNA
 A:Residues: 1243 <UPR>
 A:Cross-references: EMBL:M38698; NID:g152257; PID:g152262

S28677 Length: 243 February 11, 2000 15:52 Type: P Check: 6944 ..

1 MSLQLVGHQ SRVAAVFFA AEKQDLISLF DAQPERFTF AAQWEGDQ
 51 ARHLEVEEG TLRAVRLND GRRVITGFLR PGDLGVSVK EHYLYTEAI
 101 THVELRFSR RFESESARA PHLRQULFSR LCDMAAAD QWVLSRRA
 151 EEKVAQFLIM MARGOSENRR PYIELPMTRL DVADYLGMTI ETVSRITKL
 201 AGSGVIAIVG RHAVALIKMD ALIALADGEC DGAQRSGARY AKA

11AA_SEQUENCE 1.0
 P1:S26139 - signalling protein ampd - *Citrobacter freundii*
 C:Species: *Citrobacter freundii*
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 17-Oct-1997
 C:Accession: S26139
 R:Kopp, U.; Wiedemann, B.; Lindquist, S.; Normark, S.
 A:Description: Wildtype and mutant ampd genes of *Enterobacter cloacae* and
Citrobacter freundii.
 A:Reference number: S26138
 A:Accession: S26139
 A:Molecule type: DNA
 A:Residues: 1187 <KOP>
 A:Cross-references: EMBL:Z14002; NID:g40455; PID:g40456
 C:Genetics:
 A:Function: regulator of the production of beta-lactamase

S26139 Length: 187 February 11, 2000 15:52 Type: P Check: 4831 ..

1 MLENGWLVQ ARHVSPPHD CRPEDEKPTL LVVHNISLPP GEGGPMWIDA
 51 LFTGTIDPDA HPFPAEIAHL ALSADCLIRR DGEVQVYVF DKRAWHAGVS
 101 MYGGERCND FSIQIELEGT DTPPYDAQY EKLAVTQTL IGRYPAIADN
 151 ITGHSIDIAPE RKTDPGPAPD WSRFHMLTT SSDKEIT

11AA_SEQUENCE 1.0
 P1:A48901 - signalling protein ampd - *Enterobacter cloacae*
 C:Species: *Enterobacter cloacae*
 C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-May-1998
 C:Accession: A48901; S26138
 R:Kopp, U.; Wiedemann, B.; Lindquist, S.; Normark, S.
 A:Title: Sequences of wild-type and mutant ampd genes of *Citrobacter freundii*
 and *Enterobacter cloacae*.
 A:Reference number: A48901; MUID:93199292
 A:Accession: A48901
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1187 <KOP>
 A:Cross-references: EMBL:Z14003; NID:g40905; PID:g40906
 A:Note: sequence extracted from NCBI Backbone (NCBIN:127354, NCBIF:127356)
 C:Genetics:
 A:Gene: ampd

A48901 Length: 187 February 11, 2000 15:52 Type: P Check: 4831 ..

1 MLENGWLVQ ARHVSPPHD CRPEDEKPTL LVVHNISLPP GEGGPMWIDA

51 LFTGTIDPDA HPFPAEIAHL ALSADCLIRR DGEVQVYVF DKRAWHAGVS
 101 MYGGERCND FSIQIELEGT DTPPYDAQY EKLAVTQTL IGRYPAIADN
 151 ITGHSIDIAPE RKTDPGPAPD WSRFHMLTT SSDKEIT

11AA_SEQUENCE 1.0
 P1:S40867 - ferridoxin-NADP+ reductase (Ec 1.18.1.2) - *Escherichia coli*
 N:Alternate names: methyl viologen-resistance protein
 C:Species: *Escherichia coli*
 C:Date: 19-May-1994 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
 C:Accession: S40867; A47077; A26225; B45248; G65198; Q00073
 R:Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.
 A:Title: Analysis of the *Escherichia coli* genome. III. DNA sequence of the
 region from 87.2 to 89.2 minutes.
 A:Reference number: S40802; MUID:93347969
 A:Accession: S40867
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1248 <PRU>
 A:Cross-references: EMBL:L19201; NID:g304961; PID:g305027
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October
 1993

R:Blanch, V.; Reichard, P.; Eliasson, R.; Pontis, E.; Krook, M.; Jorvall, H.;
 Hagbard-Ljungquist, E.
 J. Bacteriol. 175: 1590-1595, 1993
 A:Title: *Escherichia coli* ferridoxin NADP+ reductase: activation of E. coli
 anaerobic ribonucleotide reduction, cloning of the gene (fpr), and
 overexpression of the protein.
 A:Reference number: A47077; MUID:93194782
 A:Accession: A47077
 A:Molecule type: DNA
 A:Residues: 143, 'A', '45-150', 'T', '152-248' <BIA>
 A:Cross-references: GB:L04757
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI Backbone (NCBIN:127345, NCBIF:127346)
 R:Moriyo, M.
 J. Bacteriol. 170, 2136-2142, 1988

A:Title: Isolation and characterization of methyl viologen-sensitive mutants of
Escherichia coli K-12.
 A:Reference number: A26225; MUID:88198002
 A:Accession: A26225
 A:Status: significant sequence differences
 A:Molecule type: DNA
 A:Cross-references: GB:M19644
 A:Note: original sequence reported in A26225 was incorrect due to frameshifts;
 carboxyl-terminal segment of glpX protein was misidentified as part of mvvA
 gene product

R:Turniger, V.; Boos, W.; Sweet, G.
 J. Bacteriol. 174: 6981-6991, 1992
 A:Title: Molecular analysis of the glpFKX regions of *Escherichia coli* and
Shigella flexneri.
 A:Reference number: A45248; MUID:93015762
 A:Accession: B45248
 A:Molecule type: DNA
 A:Residues: 1-93, 'RDSICSKKCRTRAKRYGCGWQVQRLALIRFCN', '127-130' <TRU>
 A:Note: sequence extracted from NCBI Backbone (NCBIF:117058)
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
 Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,
 J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: G65198
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-248 <BLAT>
 A:Cross-references: GB:AE000467; GB:U00096; NID:g1790356; PID:g1790359;
 UMGp_b3924
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:

Mon Feb 14 08:07:32 2000

plc.cat

Page 131

A:Gene: fpr: mvrA
A:Map position: 88 min
C:Keywords: NADP, oxidoreductase

540867 Length: 248 February 11, 2000 15:52 Type: P Check: 3905 ..

1 MADWATGKAT KYONWTDALF SLVYHAFVL PTAGOFTLKG LEIDGENVOR

51 AYGVNSPDN PDLEFFLYTV PDGKLSFRLA ALKPGDEVOV VSEAGGFVL

101 DEVPHCETIM MLAGTALGP YLSILOGRD LDREKNLVIV HAARYADLS

151 YELIMOELER RYEGKRLIQT VSKRETAGS LTRGIPALIE SGELESTIGL

201 PANKRISHVA LCGNPQWVD TOOLLKETRO MKHLRRRPG HMTAHHY

!!AA_SEQUENCE 1.0

P1:A64750 - hypothetical protein b0249 - Escherichia coli

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-May-1998

C:Accession: A64750

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: A64750

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 179 <BLAT>

A:Cross-references: GB:AE000133; GB:U00096; NID:92367099; PID:g1786443;

UMGP:b0249

A:Experimental source: strain K-12, substrain MG1655

A64750 Length: 79 February 11, 2000 15:52 Type: P Check: 8141 ..

1 MTSVLLPQG PPTRRQAAV TTYYSNTLE DDQGSHERLV VDTGGRMVM

51 RANWFEPDAG EGINRYRTS GIRTDIATR

!!AA_SEQUENCE 1.0

P1:C65059 - hypothetical protein b2775 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997

C:Accession: C65059

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: C65059

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1425 <BLAT>

A:Cross-references: GB:AE000361; GB:U00096; NID:92367160; PID:g2367161;

UMGP:b2775

A:Experimental source: strain K-12, substrain MG1655

C65059 Length: 425 February 11, 2000 15:52 Type: P Check: 1867 ..

1 MOHNSYRMT TLAISFSGG VSFDAVLYR IYDIPAKRV GFSNTEIGLI

51 MSTFGIAIIV LVAPOGVAD KESHRKMTS AMITGLLGL IMATYPIPLMV

101 MCQIOIAFAI TTILMLSVS IKAASLLGDH SEQKMGWV BGLRGVWVS

151 LAVETMWVS RFAPDDSTSL KTVIIYSV YIILGICWF EYSDNNLRIS

201 ANNEKOSFO LSDILAVLR STTWYCSMV FGVFTIYAI SYSTNYLTEM

251 YGMSIVASY MGIIVINKFR ALGGLGII TTYSKVSP RVIIQLSVLG

301 LITLITALLVT NSNPQVAMG IGLILLGLFT CVASRLIYVA CGGARPTSPY
351 INGTIVGICS VIGFLPDVEV YPIIGHMODT LPAAEAVRMWLMGMAIGM
401 VIVFTFLFQ KIRADSAPA MASSK

!!AA_SEQUENCE 1.0

P1:B65068 - hypothetical protein b2849 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 14-Nov-1997

C:Accession: B65068

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: B65068

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1145 <BLAT>

A:Cross-references: GB:AE000368; GB:U00096; NID:92367165; PID:g2367167;

UMGP:b2849

A:Experimental source: strain K-12, substrain MG1655

B65068 Length: 145 February 11, 2000 15:52 Type: P Check: 6883 ..

1 MRYTDIEFS QIHEMYMHD IVNSDSKKP RPLKRLNA ENVLQOTISM

51 TINSRYVWNV SYNVVYKSK VKNYSIRSV NDEFSLTDE INSFRETLVL

101 SSIDSLSKLV LNNLSLVLF STVRRNNRA KMWNEFDSWI CIRCC

!!AA_SEQUENCE 1.0

P1:F64963 - nicotinate-nucleotide--dimethylbenzimidazole

phosphoribosyltransferase (EC 2.4.2.21) - Escherichia coli

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 12-Feb-1999

C:Accession: F64963

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;

Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor,

J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64963

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1359 <BLAT>

A:Cross-references: GB:AE000291; GB:U00096; NID:g1788298; PID:g1788300;

UMGP:b1991

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: cobT

C:Function:

A:Description: catalyzes the reaction of beta-nicotinate D-ribonucleotide and

dimethylbenzimidazole to nicotinate and

N(1)-(5-phospho-alpha-D-riboseyl)-5,6-dimethylbenzimidazole

C:Keywords: glycosyltransferase; pentosyltransferase

F64963 Length: 359 February 11, 2000 15:52 Type: P Check: 7029 ..

1 MQLIADLNT IPADSTAMS RAQRHIDGLL KPVSLGKLE VLAIOJAGMP

51 GLNGIRHVAK KAVLYMCADH GWBEVVALS PKRYTALDAE NMIRGTTGVC

101 VLAQAGANV HVIDVGIDTA EPIPLGNR VARSGNATS APAMSROAE

151 KLIDVICT QELAKNGYTL FGVGELGMAN TTPAAIYST ITGRDEEYV

201 GIGANLPDK LANKIDVRR ATLNPQNPQ DGVDVIAKVG GFDVLGIAGY

251 MGAAACGCP VLLDGFLLSYA AALACOMSP AKEPYLPSH LSAEGARIA
 301 LSHLGLPEYL NMEKRLGEGS GAALAPPIE AACAIYNNMG ELAASNYLP
 351 GNTTSDLNS
 !!AA_SEQUENCE 1.0
 P1:T00211 - type II secretion pathway related protein etpc - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 07-May-1999
 C:Accession: T00211
 R:Blattner, F.R.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota, Y.; Yamachi, Y.; Iida, T.; Yamamoto, K.; Honda, T.; Han, C.; Ohtsuda, A.; Kasamatsu, M.; Hayashi, T.; Kuhara, S.; Shingawa, H.
 DNA Res. 5, 1-9, 1998
 A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak.
 A:Reference number: 214127; MUID:98290540
 A:Accession: T00211
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-291 <MAK>
 A:Cross-references: EMBL:AB011549; NID:d1204561; PID:d1032719
 A:Experimental source: strain EHCC O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: etpc
 A:Genome: plasmid p0157

T00211 Length: 291 February 11, 2000 15:52 Type: P Check: 5937

1 MLEFLSRGD KGLFKDIYL KMLTPNRLC VILLIAGYQL VSVIHFMLT
 51 QAAVPEQLSR VSAPETAVTG DQEEERFVET LFGRAPISS EGRAOETMPS
 101 LSDDLASGED LVRGILYSS VAESVAIFA HNNROFSLV GEKPSYDAT
 151 ISAFSDHIY INYQKTVSL PLRYDNTERR NAYDNNMLTV GDVITODNFR
 201 VESVFDMISF SAVTVNNTLS GYRLIPGKS SLFYNAGLMD NDLAVSNGS
 251 ELRDTROAQO IMKQPELKE IKITVERDQ LYDAIIVAGE N
 !!AA_SEQUENCE 1.0
 P1:H64726 - yabp protein - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 15-May-1998
 C:Accession: H64726
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: H64726
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-316 <BLAT>
 A:Cross-references: GB:AE000116; GB:U00096; NID:g1786240; PID:g1786242;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: yabp

H64726 Length: 316 February 11, 2000 15:52 Type: P Check: 98

1 MKVSVGPVP TLLNMSKNDI YKMGSGDKMD VKMNIQRLM ETLRLFLMSD
 51 KQTEAYKLIF NFVNNOTGNI NASEFTGAI NENEREKFIN SLELFNKAT
 101 CAKNDELIVA KGNMRYVAQT FGDIELSVTF FIKKNICIQ TLQHKQGN

151 LGVDRKAVL PGYDMRDCYL GKRTMGKSD ILYERPGWNA NGVLPRTL
 201 PRVIALRIVL TWIYLP
 !!AA_SEQUENCE 1.0
 P1:A64862 - ycgZ protein - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 13-Sep-1998
 C:Accession: A64862
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: A64862
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-78 <BLAT>
 A:Cross-references: GB:AE000215; GB:U00096; NID:g1787405; PID:g1787411;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ycgZ

A64862 Length: 78 February 11, 2000 15:52 Type: P Check: 1943

1 MHQNSVTLDS AGAIRRYFAK ANLHTQOETL GEIYTEILKD GRNLSKRSLS
 51 AKLLCRLEHA TGEEOKHYN ALIGLIFE
 !!AA_SEQUENCE 1.0
 P1:A64882 - ycgI protein - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 24-Oct-1998
 C:Accession: A64882
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: A64882
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-262 <BLAT>
 A:Cross-references: GB:AE000230; GB:U00096; NID:g1787578; PID:g1787586;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ycgI

A64882 Length: 262 February 11, 2000 15:52 Type: P Check: 6021

1 MIIRICITTL PYLPEPENT MIVTPRAER GAPPGETEHY GRSLGAPLI
 51 WFPAPASRE SGLLIAGTIG DENSSVYTL CALRTITPSL RRRHVVLGVN
 101 PDCGLGARA NANGVLDNRN FPAANKKEGE TYRWNSSAE ERDVLVLTGD
 151 KFGSEPTQA LCOLIHRIOF AMVVSFDPL ACTIEDPRASE LGEWLAQAFE
 201 LPVTSVGE TPQSGFSWCA DLNLICITAE FPPISSDEAS EKYLFAMANT
 251 LRHFKDAIR PS

!!AA_SEQUENCE 1.0
 P1:S20452 - hypothetical protein X - Klebsiella pneumoniae (fragment)
 C:Species: Klebsiella pneumoniae
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
 C:Accession: S20452; S21837
 R:Meulenbergh, J.J.M.; Sellink, E.; Rieggman, N.H.; Postma, P.W.
 Mol. Gen. genet. 232, 284-294, 1992

Mon Feb 14 08:07:32 2000

pir.cat

Page 133

A:Title: Nucleotide sequence and structure of the Klebsiella pneumoniae pgq operon.
A:Reference number: S20452; MID:92212293
A:Accession: S20452
A:Molecule type: DNA
A:Residues: 1-271 <MBD>
A:Cross-references: EMBL:X58778; NID:943903; PID:943904
S20452 Length: 271 February 11, 2000 15:52 Type: P Check: 9681 ..

1 SVPGDDITIL WQOLAIKOL IAHSGRLRL CLSADIERC REDVILAVYA
51 HIEGAGGFDG EGRDLQAFYA AGVRSIGPFM NIANREGSGV NSPFGSPDI
101 GGILTAAGID LIKOVNALKM QIDVSHMNEK AFMDTAHAT SPLVATHSNA
151 HALCPQPNL TDQGLRAIRD SGGVGVNFG NAFIRADGR DSDTLLITIV
201 RHIDYLNIM GEDHVALGSD FPGITLPEEL GDVAGLPRLI NTLRASGYDO
251 LVLDKILWEN WLRYLKNWQ Q

!!A:SEQUENCE 1.0
P1:S01838 - nifH protein - Klebsiella pneumoniae
C:Species: Klebsiella pneumoniae
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 09-Sep-1997
C:Accession: S01838; S01703
R:Arnold, W.; Rump, A.; Klipp, W.; Pfeifer, U.B.; Puchler, A.
J. Mol. Biol. 203, 715-738, 1988
A:Title: Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae.
A:Reference number: S01836; MID:89094839
A:Accession: S01838
A:Molecule type: DNA
A:Residues: 1-220 <ARN>
A:Cross-references: EMBL:X13303; NID:943820; PID:943826
R:Beignon, J.; Cannon, M.; Buchanan-Wollaston, V.; Ally, A.; Setterquist, R.; Dean, D.; Cannon, F.
Nucleic Acids Res. 16, 9860, 1988
A:Title: The nucleotide sequence of the nifH, nifY, nifX and nifW genes of K. pneumoniae.
A:Reference number: S01702; MID:89041575
A:Accession: S01703
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-169, 171-217, 'GLTGITLWKNAFNCFVLVAR' <BEY>
A:Cross-references: EMBL:X12595
C:Genetics:
A:Gene: nifY

S01838 Length: 220 February 11, 2000 15:52 Type: P Check: 7149 ..

1 MSDNDLFWR MLAFOSLPD LOPAOIVDWL ADESGETLP ERLATLPOQ
51 LAASFPSATA VMSPARMSRV MASLOGALPA HLRIVPAOR TPQLIAFCS
101 QDGLVINGH GCGRLFEIYA FDEQGWILD LRRYPAPHQ OEANEVRARL
151 IEDQQLLCQ EIGGPAAR IHRIRHPKA QPGTTIOAC EAINTLIAGR
201 LPPWLAKRLN RDNPLERVE

!!A:SEQUENCE 1.0
P1:S22619 - hypothetical protein - Salmonella choleraesuis
C:Species: Salmonella choleraesuis
C:Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 09-Sep-1997
C:Accession: S22619
R:Brown, P.R.; Komana, T.K.; Reeves, P.R.
Mol. Microbiol. 6, 1385-1394, 1992
A:Title: Molecular analysis of the rfb gene cluster of Salmonella serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B.
A:Reference number: S22613; MID:92349966

A:Accession: S22619
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-336 <BRO>
A:Cross-references: EMBL:X61917; NID:947004; PID:947011
A>Note: the nucleotide sequence was submitted to the EMBL Data Library. //
S22619 Length: 336 February 11, 2000 15:52 Type: P Check: 7734 ..

1 MNKKVYLDI SMSNKGIGR FIDEISKRLC DISKEELYR CASPLAPGL
51 AVNIFLRKKT DVPFLPGIYR PLFCCKRPII TIHDLNHLI DNSSLEFRL
101 EYNFIRKQC RKAKITTS NFSKRIYAN SGVNPKIYI YNGVSSLFN
151 ADVKPLNIGY KYLLCVGNKR THKNECVIS AFAKADIPS IKLVGTGIPC
201 NDLEKLIQH GLSERVKEFG EVSEKDLPSL YKSLGLVFP SLXEGFGLPV
251 VEGMAGIYV LTLTSLPE VAGDAAILVD PLSEDAITKG ISRLNDBEL
301 RKHLIQKGLL RAKRFNNQNV VSEIEWVTE AGDGNK

!!A:SEQUENCE 1.0
P1:A53302 - hemF 5'-region hypothetical protein - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 09-Sep-1997
C:Accession: A53302
R:Xu, K.; Elliott, T.
J. Bacteriol. 175, 4990-4999, 1993
A:Title: An oxygen-dependent coproporphyrinogen oxidase encoded by the hemF gene of Salmonella typhimurium.
A:Reference number: A53302; MID:93352403
A:Accession: A53302
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-289 <XU>
A:Cross-references: GB:U19503; NID:9310640; PID:9310641

A53302 Length: 289 February 11, 2000 15:52 Type: P Check: 5663 ..

1 MSTRKLKTL TSRROYLKTG LALITLSGMS HAAVAEETLK TSNQSKPKT
51 KKTGSKRLVM LDPGHGIDT GAIGRNGSQ KHVLAIAIKN VRATLRNHGI
101 DARLRTGDT FIPLYDRVEI AHKAGADLPM SIHADGETNP KAAGSVFAL
151 SNRGAASSMA KYLSERENRA DEVAGKRATD RDLHLOQVLF DYVOTDLIKN
201 SLTIGSHILK KIRPIKILHS RTEQAALFV LKSPSIPSVL VETSFITNPE
251 EERLIGTTAF ROKIATAIAN GIISYFWFD NOKAHTTKR

!!A:SEQUENCE 1.0
P1:S23906 - hypothetical protein 1 - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
C:Accession: S23906
R:Truniger, V.; Sweet, G.; Boos, W.
Submitted to the EMBL data library, March 1992
A:Description: glpX, a new glp gene is located at the end of the glpFK operon of Escherichia coli
A:Reference number: S23905
A:Accession: S23906
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135 <TRU>
A:Cross-references: EMBL:Z11766; NID:947027; PID:947029

S23906 Length: 135 February 11, 2000 15:52 Type: P Check: 4054 ..

1 MADWTGKVT KVNWTDALF SLTVHAPVLP FTAQGFITKG LELDERVOR

```
51 ASYVNSPDN PDLEFYLVTV PDGKISPLRA ALKPDDEVYV VSEQLASLFW
101 MKYRTAKRYG CWOPVORLAT IYFCNIGKD LDRFK

!!AA_SEQUENCE 1.0
P1:T14653 - histone H5 like protein - Yersinia pestis plasmid pMT1
C:Species: Yersinia pestis
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14653
R:Hu, P.; Elliott, J.; McCreedy, P.; Skowronski, E.; Garnes, J.; Kobayashi, A.;
Carrano, A.V.; Brubaker, R.; Garcia, E.
submitted to the EMBL Data Library, March 1998
A:Description: Structural organization of virulence determinants in three
Yersinia pestis plasmids.
A:Reference number: 218168
A:Accession: T14653
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-99 <HUP>
A:Cross-references: EMBL:AF053947; NID:g2996286; PID:g2996296; PIDN:AAC13176.1
C:Genetics:
A:Gene: Plasmid pMT1

T14653 Length: 99 February 11, 2000 15:52 Type: P Check: 2204 ..

1 MRATYEGKN VSPSAVYSR NDLRADIVE TAAPIVPMK RGTAEVENAKP
51 VQRTFEELE STADEGAG LKQIGNOIGV KAKGIVEMTE GILKAGGE

!!AA_SEQUENCE 1.0
P1:T14952 - hypothetical protein - Yersinia pestis plasmid pMT-1
C:Species: Yersinia pestis
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14952
R:Findley, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.
Infect. Immun. 66, 5731-5742, 1998
A:Title: Complete DNA sequence and detailed analysis of the Yersinia pestis
KRM5 plasmid encoding murine toxin and capsular antigen.
A:Reference number: Z18268; MUID:99043898
A:Accession: T14952
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-144 <LIN>
A:Cross-references: EMBL:AF074611; NID:g3883003; PID:g3883036; PIDN:AAC82696.1
C:Genetics:
A:Gene: Y1036
A:Genome: Plasmid pMT-1

T14952 Length: 144 February 11, 2000 15:52 Type: P Check: 2530 ..

1 MKPAIRLLE POFLYTGIL CGIQFPGIS VAEIRFIDQ RITASKRATT
51 VEGKVVPSA AYSSRDULTA DDIETAAPD IVPKRGTAE VEAKPVQRTF
101 REELSIADC EGTAGLRQIG NOIGYAKKI VEMTEGILKA OGGE

!!AA_SEQUENCE 1.0
P1:S70883 - hypothetical protein 3 - Vibrio cholerae
C:Species: Vibrio cholerae
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 17-Mar-1999
C:Accession: S70883
R:Comstock, L.E.; Johnson, J.A.; Michalski, J.M.; Morris Jr., J.G.; Kaper, J.B.
Mol. Microbiol. 19, 815-826, 1996
A:Title: Cloning and sequence of a region encoding a surface polysaccharide of
Vibrio cholerae O139 and characterization of the insertion site in the
chromosome of Vibrio cholerae O1.
A:Reference number: S70876; MUID:96417860
A:Accession: S70883
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-227 <COM>
A:Cross-references: EMBL:U47057
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January

1996
S70883 Length: 227 February 11, 2000 15:52 Type: P Check: 2492 ..

1 MARLFIWNL SRVKKRIISV LVDAFFIFS FYSAYWVRG NVEIHSOSI
51 PYLLATIVV TISFTRLG YRAIRLYTE HALAVSVGT LISASVALA
101 AFYDAPVPR SLEIITGFL CLICGSRLI VRLVSGLNG KKKRVLLYG
151 AGSAGROLAI ALRNSENVYR AGFIDNDKTL ENTVMGMQV HDVSNAAVLV
201 DKYDVTOILL AVPSARALS SDCAKI

!!AA_SEQUENCE 1.0
P1:S54442 - membrane associated protein 17.9K - Vibrio cholerae
C:Species: Vibrio cholerae
C>Date: 06-Sep-1996 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997
C:Accession: S54442
R:Johnson, G.; Lebens, M.; Holmgren, J.
Mol. Microbiol. 13, 109-118, 1994
A:Title: Cloning and sequencing of Vibrio cholerae mannose-sensitive
haemagglutinin pilin gene: localization of mshA within a cluster of type 4
pilin genes.
A:Reference number: S54441; MUID:95075288
A:Accession: S54442
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-150 <COM>
A:Cross-references: EMBL:X77217; NID:g496755; PID:g496757
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January
1994

S54442 Length: 150 February 11, 2000 15:52 Type: P Check: 6509 ..

1 MKIGERVRF VLMFVLVYL LTFMSVWRS MFSDLHTAL EMTRLOIDR
51 ANTYQEWTL QGRPALLOIE QAEIPMOHGM VFPKIDQVD CERVELFLVP
101 DRKVIDWLR VTDLORANGY OCRVQGRDV QPDVELKRVY FAIINAFLWR

!!AA_SEQUENCE 1.0
P1:G64102 - phosphatidate cytidylyltransferase homolog - Haemophilus influenzae
(strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 24-Oct-1997
C:Accession: G64102
R:Flaischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
Kerlavage, A.R.; Buit, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips,
C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.;
Nguyen, D.T.; Sauder, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.;
Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
Venter, J.C.
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd
A:Reference number: A64000; MUID:95350630
A:Accession: G64102
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-288 <TIGR>
A:Cross-references: GB:U32773; GB:L42023; NID:g1573932; PID:g1573940;
TIGR:H1019

G64102 Length: 288 February 11, 2000 15:52 Type: P Check: 8973 ..

1 MKQRYVLSAI VLIAAVLCAL FLTFPEYFAL ALGAVAIIGI WENQFARLK
51 QPLIRFEVTT FLGVIFLML YTEGNYLDAG RYFEQHLQLL LINAVSWMGL
```

101 ALLVIVSYR SAKFWSKNPL LQLLAFSTL IPFVAGVRL RLEHYTHDPY
 151 HGLLELLVYF ILVMADSGA YFSGRAGKR KLAPKVSPEK SMEGVIGLI
 201 TALVIALEFI HESNNTLVGD ENINGFILLS VATVAISVIG DLESMPKRE
 251 SGVKSQSLI PGHGVLDRI DSIYAVPFF SYFFEFVL

!!AA_SEQUENCE 1.0
 P1:H64157 - sufl protein homolog HI0733 - Haemophilus influenzae (strain Rd
 KW20)
 C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Oct-1997
 C:Accession: H64157
 R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;
 Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;
 McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
 Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Melman, J.R.; Hanna, M.C.;
 C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Uterback, J.R.; Hanna, M.C.;
 Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.;
 Fuhmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;
 Venter, J.C.
 A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd
 A:Reference number: A64000; MUID:95350630
 A:Accession: H64157
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1311 <DIGR>
 A:Cross-references: GB:U32756; GB:I42023; NID:91573729; PID:91573736;
 TIDR:HI0733

H64157 Length: 311 February 11, 2000 15:52 Type: P Check: 4

1 MPRLSRQQL KTAIASTALS TVPAPLLAS REKLVPPLI EVRGRREIVL
 51 TMOETNPPLD GSHNVTWGF NGNYLGPTEK IKSGSAFLN YHNNLPQSA
 101 LSIQGLQASG ELPGGARVYL KKGSMAPIV PIEOPASQW YRSATLANSA
 151 YQYRGLAGM WLEDEQSLK ANLPNKYVD DIPILIDME FNNQGLDLFK
 201 QNOPHFVGNR LLYNGIEAPY LDVARGWIRL RLINSLARA YDLALDDQE
 251 MLINADLGF LPRKSVKSL VLSPEERAEI LVNMKLTLY LSLAEVAVAC
 301 TENKRYVLR R

!!AA_SEQUENCE 1.0
 P1:H71692 - hypothetical protein RP358 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 07-May-1999
 C:Accession: H71692
 R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheit-Ponten, T.;
 R. Anderson, S.G.E.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.;
 Kurland, C.G.
 Nature 396, 133-140, 1998
 A>Title: The genome sequence of Rickettsia prowazekii and the origin of
 mitochondria
 A:Reference number: A71630; MUID:99099499
 A:Accession: H71692
 A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-405 <AND>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:93860788; PID:el342662;
 PID:93860918
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: RP358

H71692 Length: 405 February 11, 2000 15:52 Type: P Check: 1607

1 MOYLILSLIF LIPSLGMLTG LSIATVAEP LLSIIITGFI SFIOKQEPNK
 51 KFFIKLPRE GFDDIFCIIN KIKTELEFTA WCFISCLAV HPINSLVTEF
 101 KVFYLLPFR VSNATVPOV LYIKNSLILG IITAILLEFI EYSSGFLR
 151 MEKTHFLGYM LDRGCALLSI TTVAITILF SNGHINSEFI LYIVLYLIS
 201 ISDSLAFSLG FSIQGIIFIL ARLKITEFK LTIISLITGS LLEPVIARQI
 251 DPQNLSEKYL ATOPSAARL FIMHFVANKI IIRPIGYGF ASSXYIEFGD
 301 NAMIDYRGEK LHPPLPHPN NILQITTELG ILGLAFILC VYKIKLEIDN
 351 IKVSNFRAS YSCFINTYII GMISINWQT WMLISGIVL VLMLLVKPD
 401 IIIDN

!!AA_SEQUENCE 1.0
 P1:E72060 - c1470 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
 C:Species: Chlamydia pneumoniae
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999
 C:Accession: E72060
 R:Kallan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
 Grimwood, J.; Davis, R.W.; Stephens, R.S.
 Nature Genet. 21, 385-389, 1999
 A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: E72060
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-252 <ARN>
 A:Cross-references: GB:AE001643; GB:AE001363; NID:94376876; PID:94376883
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: Cpn0589

E72060 Length: 252 February 11, 2000 15:52 Type: P Check: 2665

1 MOICVTGVVL RSRPLGRNHT LITLTFPEGL FTFPAKQGT LQCDYRETLV
 51 PISLQKTLH RNSGRPLKLT HGDLNLFEN IKQYVALLA SGKMIOALLA
 101 SQWKEKPSHK LPSLFLNPLH RIPPSSNPF FPAIVYKLL QYSGILDLTP
 151 ACSLCKASLP YACRYQGHK LCKKHQKQA ISIEKEEEOI LQALIHAKQF
 201 SELLAIAPF IAIKRIEYL FDSIQEKKK ERNSSEDPYH EILRLKVVH
 251 PY

!!AA_SEQUENCE 1.0
 P1:E72031 - c1578 hypothetical protein - Chlamydia pneumoniae (strain CWL029)
 C:Species: Chlamydia pneumoniae
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Apr-1999
 C:Accession: E72031
 R:Kallan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;
 Grimwood, J.; Davis, R.W.; Stephens, R.S.
 Nature Genet. 21, 385-389, 1999
 A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: E72031
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <ARN>
 A:Cross-references: GB:AE001643; GB:AE001363; NID:94377118; PID:94377121
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: Cpn0809

E72031 Length: 493 February 11, 2000 15:52 Type: P Check: 6017

```
1 MSISSSGCD NQKINMSOVL TSTPQGVPOO DKLSCNFKQ IQOTRCKNT
51 EKHSDATLNG ASGRDKTSST TKETAPQOQ VAAKSESSES OKKADTGV5
101 GAAATTAASNT ATKIAMOTSI EEASKSMEST LESLOSLSAA OKMEVEAVV
151 AALSCKSSGS AKLETPPELPK PGVTPRSEVI EIGLALAKAI QTIGEARISA
201 LSNVASTQAO ADQTNKLGLE KQAIKIDKER EBYQEMKAAE OKSKDEGTM
251 DTVNTVMIV SVAITVISIV AAFITCGAGL AGLAAGAVG AAAAGNAGA
301 AAATVATQI TVQAVVQAVK QAVITAVROA ITAIAKAVK SGIAKIFITL
351 VKAIAKAIK GISKVFAKGT OMTAKNPKL SKVISSLSK WVTYGVGVV
401 AAPALGKIM QMOLEMQON VAOFOKEVGR LOAADMISM FTQFWQOASK
451 TSKOTGESN EMTOKATKLG AQILKAYAI SGATGAHRT NNF

!!AA_SEQUENCE 1.0
P1:S61492 - hypothetical protein 2 - Chlamydia psittaci
C:Species: Chlamydia psittaci
C>Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 09-Sep-1997
C:Accession: S61492
R:Rockey, D.D.; Heinzen, R.A.; Hackstadt, T.
Mol. Microbiol. 15, 617-626, 1995
A:Title: Cloning and characterization of a Chlamydia psittaci gene coding for a
protein localized in the inclusion membrane of infected cells.
A:Reference number: S61491; MUID:95302975
A:Accession: S61492
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-270 <RCS>
A:Cross-references: EMBL:L35036; NID:9516598; PID:9516600

S61492 Length: 270 February 11, 2000 15:52 Type: P Check: 5418 ..

1 MSRYSSNKKH HSRKTERNS TSWEPIAEDY HKIVQEGHY YKVEVILPKL
51 LPLMLQSED SLVDIGCGQG ILERAIPKEC GYLGLDISPS LISTARKLR
101 SRDHEKIOD LTKRLVLETP QSFSAVAIL SLQNNETPER AIKNTSKLLN
151 DGRFFMYLN HPCFRIPRVS SMHYDEDKL LSRKIDRYLS KITVSIYANP
201 GKQSSSSIS FHFPLSYWTQ ALSKYGFVIE NMEWISPKK SIGAHAKAEN
251 LCREPEPLFL MISCIKTHKN

!!AA_SEQUENCE 1.0
P1:JC5205 - sulfur-rich protein - Chlamydia psittaci
C:Species: Chlamydia psittaci
C>Date: 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 07-May-1999
C:Accession: JC5205
R:Hsia, R.; Bayoili, P.M.
Gene 176, 155-162, 1996
A:Title: Sequence analysis of the omp2 region of Chlamydia psittaci strain
GPIC: Structural and functional implications.
A:Reference number: JC5203; MUID:97075924
A:Accession: JC5205
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <HSI>
A:Cross-references: GB:U41759; NID:91783376; PID:91783383
A:Experimental source: strain GPIC
C:Genetics:
A:Gene: srp

JC5205 Length: 160 February 11, 2000 15:52 Type: P Check: 8587 ..

1 MLTGVENSES GVIDLIKPGI DDVAKNETVO VTLVNSVLGW CKAHIYDPIK
```

```
51 TSKIVQSAF QIMVUVGI LLIALGLTF VLQGLQKNA FLFLIYAVIC
101 LVKLITTSVF MEKPTPEKX RICKRLIATF EDIIDDQIN QSNITTTES
151 SDVINTAFOS

!!AA_SEQUENCE 1.0
P1:A71529 - hypothetical protein CT324 - Chlamydia trachomatis (serotype D,
strain UM3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
C:Accession: A71529
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;
Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.
A:Reference number: A71570; MUID:99000809
A:Accession: A71529
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <ARN>
A:Cross-references: GB:AE001273; GB:AE001273; NID:93328737; PID:93328742
A:Experimental source: serotype D, strain UM-3/Cx
C:Genetics:
A:Gene: CT324

A71529 Length: 303 February 11, 2000 15:52 Type: P Check: 4189 ..

1 MYKAPHPHS SISGALPLHS SRGAYHDSL GDLSPISLT PRALSVLTPB
51 PAOSKRVQA VANTIGDFLQ RMKRLHLYI LWSGLILICH SSVALVLSIW
101 LGIGGAGV LGISANFLD KENKYPHLNS LMNTINNGLO QUDPNTROV
151 LLATVIASIS ALIYASPAI GFILGAFGH QTSLLAYGY RFKSGGYIA
201 DQELDKQEK RIRQAIMQR LIRNOMILOR RLDPLQKRAH RPOQSCITTI
251 RPLNRIISLP EMEHEILK KPOVVAMUD NKIAQLNOSL VOLYDRPTV
301 IEQ

!!AA_SEQUENCE 1.0
P1:C71569 - probable ribonuclease HII - Chlamydia trachomatis (serotype D,
strain UM3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Nov-1998
C:Accession: C71569
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.;
Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.
A:Reference number: A71570; MUID:99000809
A:Accession: C71569
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <ARN>
A:Cross-references: GB:AE001273; GB:AE001273; NID:93328388; PID:93328396
A:Experimental source: serotype D, strain UM-3/Cx
C:Genetics:
A:Gene: rnhB_1

C71569 Length: 300 February 11, 2000 15:52 Type: P Check: 1406 ..

1 MSSFVSQLS PSLESLREQ LEKKGFTISI PPHVTFQGRS PTVSCVYGS
51 KTIYQKGT QEFYEFLEP ELQTFSSON VOQDLRSRIG VDSGCGDF
101 GPLCTAGYA SSPAIEALY KTSICDSKLI PDAIILSLAQ NITSLCAKY
151 TLPEPKINA LYANFONLNS LIAWTHATII DNLAHPHAGA VFAISDPFAS
```

201 SERVULQAVR KKCSDEILIO RHRAEDOVVY AASILAREA FLSSIHAALES
251 OYORILIKGA SGRVORAKE ILHNGOVVL EKVCTHFTI FNEVLGSGNQ

11AA_SEQUENCE 1.0
P1:B53203 - hypothetical protein 2 - Desulfovibrio vulgaris (strain Miyazaki)
C:Species: Desulfovibrio vulgaris
C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 09-Sep-1997
C:Accession: B53203
R:Kittamura, M.; Kojima, S.; Ogasawara, K.; Nakaya, T.; Sagara, T.; Niki, K.; Mura, K.; Akutsu, H.; Kumagai, I.
J. Biol. Chem. 269, 5566-5573, 1994
A:Title: Novel FMN-binding protein from Desulfovibrio vulgaris (Miyazaki F).
Cloning and expression of its gene in Escherichia coli.
A:Reference number: A53203; MIM:94164898
A:Accession: B53203
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-43 <KIT>
A:Cross-references: GB:D21804; NID:g425735; PID:d1005361; PID:g476038
B53203 Length: 43 February 11, 2000 15:52 Type: P Check: 1180

1 MTVGLANPO ANPDRHARY ARLADPAGPA SRSLPDPWP RAA

11AA_SEQUENCE 1.0
P1:S17812 - hypothetical protein 192 - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Sep-1997
C:Accession: S17812
R:Burke, D.H.; Albert, M.; Armstrong, G.A.; Hearst, J.E.
Submitted to the EMBL Data Library, November 1991
A:Description: The complete nucleotide sequence of the 46 kb photosynthesis gene cluster of Rhodobacter capsulatus.
A:Reference number: S17803
A:Accession: S17812
A:Molecule type: DNA
A:Residues: 1-192 <EMB>
A:Cross-references: EMBL:Z11165; NID:g46097; PID:g46111
S17812 Length: 192 February 11, 2000 15:52 Type: P Check: 2127

1 MDLLEDFERA AHVPEEMAT HYIPEARQI GAAMSDRIG FAQVTAISR
51 LQELHALOT LVTADSVCA NGATVILIV PEGHTIGAL TVAMELRRG
101 VSVRIYFAPG LSDLSRLMAT TREDALITV GSDRVEICA KIVKTISSLT
151 KGRKRVAGG AIVSORAEL ARTGADLVIN DLSVISEFS LV

11AA_SEQUENCE 1.0
P1:S22631 - petp protein - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
C:Accession: S22631; S21001
R:Tokito, M.K.; Daidal, F.
Mol. Microbiol. 6, 1645-1654, 1992
A:Title: petp, located upstream of the fbcFBC operon encoding the cytochrome bc(1) complex, is homologous to bacterial response regulators and necessary for photosynthetic and respiratory growth of Rhodobacter capsulatus.
A:Reference number: S22631; MIM:92356828
A:Accession: S22631
A:Molecule type: DNA
A:Residues: 1-166 <TKO>
A:Cross-references: EMBL:Z12113; NID:g49287; PID:e49248; PID:g1333802
C:Keywords: DNA binding; transcription regulation
S22631 Length: 166 February 11, 2000 15:52 Type: P Check: 1875

1 MADTGAPGE TLFITDEQL KGIEMAFPA YRGFTADPDR IIDQDIYGA

51 HHRAHFINR EPGVLTTLI SVLGVTQSL NREVLTLIDD GLVESRGR
101 DKRRHILMT EKVGLIEREL SEAGRVMPA AVRAQPOAV AGFROYLEAM
151 MDPAMRRHQ MLKDAE

11AA_SEQUENCE 1.0
P1:B58883 - transcription negative regulator Chr - Rhodobacter sphaeroides
N:Alternate names: cobalt-heme resistance protein chr
C:Species: Rhodobacter sphaeroides
C>Date: 16-Oct-1998 #sequence_revision 16-Oct-1998 #text_change 16-Oct-1998
C:Accession: B58883; A56261
R:Donohue, T.J.; Newman, J.
Submitted to Genbank, November 1996
A:Reference number: A58883
A:Accession: B58883
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-213 <DON>
A:Cross-references: GB:U1283; NID:g1654106; PID:g1654107
A:Note: sequence revision of A56261 into two ORFs. This revision submission is not cited in Genbank entry RSU11283, release 106.0
R:Schlike, B.A.; Donohue, T.J.
J. Bacteriol. 177, 1929-1937, 1995
A:Title: Chr positively regulates transcription of the Rhodobacter sphaeroides cytochrome c-2 gene.
A:Reference number: A56261; MIM:95238260
A:Accession: A56261
A:Molecule type: DNA
A:Residues: 1-213 <SCH>
A:Cross-references: GB:U1283; NID:g1654106
A:Comment: This negative regulator of transcription inactivates the upstream encoded RpoE (see PIR:A58883).
C:Keywords: C:Gene: Chr
C:Keywords: transcription factor
B58883 Length: 213 February 11, 2000 15:52 Type: P Check: 818

1 MTRHVSDA LITVAACTL SEAFSLVAT HSLCDECRA RAGALDAVGS
51 SIMEETAPVA LSEGLASVM AQDROTORP APARADPRA PABLDYGR
101 RLEDVRWRL GGGVROATLP TGGBAIARLI WIPGQANPD HGRIGLETL
151 VLOGAFRDET DRFGAGDIEI ADELEHTPV AERGIQICL AATDAPLREN

201 SFLPKVQPF FRI

11AA_SEQUENCE 1.0
P1:S36981 - hypothetical protein 7 (atpC 3' region) - Synechococcus sp. (PCC 6716)
C:Species: Synechococcus sp.
A:Variety: PCC 6716
C>Date: 22-Jan-1994 #sequence_revision 12-Apr-1996 #text_change 07-May-1999
C:Accession: S36981; S31905
R:van Walraven, H.S.; Lutter, R.; Walker, J.E.
Biochem. J. 294, 239-251, 1993
A:Title: Organization and sequences of genes for the subunits of ATP synthase in the thermophilic cyanobacterium Synechococcus 6716.
A:Reference number: S36980; MIM:93371369
A:Accession: S36981
A:Molecule type: DNA
A:Residues: 1-201 <VAN>
A:Cross-references: EMBL:X70433; NID:g49234; PID:g49237
A:Experimental source: PCC 6716
S36981 Length: 201 February 11, 2000 15:52 Type: P Check: 4069

1 MYALOPRGY EIQILAPCGS RLPVPAIEE IPGALQVPAQ HORRDPITV
 51 PMSNVLANLW ARVROLOQGY DLIVFAVDM LPFLYFPWLS RYVAHLYSMA
 101 SISEVMDQAI ATVIDQYPS IGVYTRQQA TFPFGRCVC LSGGLSLY
 151 EFCADPDVL CMLGRAPKX GLEDAVAAN VTRTPKING QMDDVDYMR
 201 I

!!AA_SEQUENCE 1.0
 P1:S74634 - heme binding protein precursor ycf10 - *Synechocystis* sp. (strain PCC 6803)
 N:Alternate names: hypothetical protein s111685
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
 C:Accession: S74634
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitani, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MID:97061201
 A:Accession: S74634
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1593 <RAN>
 A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PID:d1017519; PID:g1516593
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Gene: ycf10

S74634 Length: 393 February 11, 2000 15:52 Type: P Check: 2111
 1 MLPKRTGSP NRRFHWLVK QALALDNAQ VSAQAISIE KTYFDGKRA
 51 PDSORGWNTY NYKSTLERE LKIKFNLR FNAITNLNNS EDPQIEAET
 101 IVKDAIEKI IAKYRSEDL VEDSOSAE INNSPKRSN FNIQRDLA
 151 ODEKRIIOLN RNLRQRIA TRFLIVLFI PLTVQILITN LYFAPLVQHF
 201 RVDIVAMEKI HYOETIEHY FEEFARYKET LEIKOLSEN OPLNQEKIHO
 251 ELKKAREELI ROATNSQOG IVNLADIAG LVAFVULIY FNGKSLITQO
 301 YLSQSFALN DITKVFIFIL LIDMFVGHHS AHGMEVYLEN LSHRGLEPEN
 351 RHAVYIFIAI VEVFLDSLFR LLIFNYITRQ SPISVAILEK MQO

!!AA_SEQUENCE 1.0
 P1:S76728 - hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
 C:Accession: S76728
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitani, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MID:97061201
 A:Accession: S76728
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1187 <RAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019373; PID:g1653729
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG

S76728 Length: 187 February 11, 2000 15:52 Type: P Check: 9759
 1 MLKRVQIIS GVLAAALG FTRPAQAEPT QKGNILISQ LPGRHMTLPD
 51 PEQMSFDGLA IGRVKGKVS IIQVELLPYG DEPGYEVSD NSRIYHMDGA
 101 GAEPDQDVL LQPFDDNGN YIRDEYSAA HPTWITRLDL KEVODVITSE
 151 INFQSEPVY LPVTRSPA PVARAPAPAP MPIRGLW

!!AA_SEQUENCE 1.0
 P1:S76937 - hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
 C:Accession: S76937
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitani, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
 A:Reference number: S74322; MID:97061201
 A:Accession: S76937
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1312 <RAN>
 A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PID:d1019582; PID:g1653939
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG

S76937 Length: 312 February 11, 2000 15:52 Type: P Check: 9556
 1 MIMKILITGG GGYIGSVLTP TLLAGYQVT VIDNPFQON SLAECQYET
 51 FNVIRGDCRK EDLKEQKLT ADVIIPAL VQAPICSRQ IGTITNOA
 101 VEMICQASP QQRILMPYTN SGYIGSKRK FCTESPLRP ISLYGVKVE
 151 AEKAVILARQ SMTEPLATV GMAPRMAYDL LVNDVYRAF YDRAVVIPEG
 201 HFRKRYIHR DVAKYFLHGL ENFESKAKP YVVGLEDANL SKTELCAEIR
 251 KLYPNFVILE APIGEDPKR DIYVSNRII STGETPDMSL GRGIDELIKG
 301 YTLRNSVYS NV

!!AA_SEQUENCE 1.0
 P1:S74351 - hypothetical protein s110071 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
 C:Accession: S74351
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugitani, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium

Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201
A:Accession: S74351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <K&N>
A:Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PID:d1010920; PID:g1001128
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74351 Length: 333 February 11, 2000 15:52 Type: P Check: 4892 ..

1 MRKLYFLPG LGKRFVCGGL WAEKTIKLY QECPATIV YKORENTLE
51 LDDILSQDL GOVIEVISWG FDIPIAKKL QGYRITAH STGYGDLPT
101 ALPIVCISRN TMGYWGRSP HSLIYLPNH ISEDFSLGL VADIDVLVA
151 RKSHYLKO LIPALSOQN VKVIDSYED LTGIFNRKY YLDSAEYA
201 QGVSEGFGL CPENMACGC HVESSVNGI SDYLDGFCN EKIAGYALDY
251 DQRIIGVYK GSRQWRVDP NLAEYRREN ILERLVTLL DINEFDYHP
301 SFSOTIEPL QRIAMLEK SLTKYKNE KSS

IIA_SEQUENCE 1.0
PI:S74733 - hypothetical protein sl10931 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
A:Accession: S74733
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Mura, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S74733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-156 <K&N>
A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017617; PID:g1651958
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74733 Length: 156 February 11, 2000 15:52 Type: P Check: 376 ..

1 MASTKILCK TFIISGLAF LGMTVPMLT VLRGNANOT GPSQGTPOOP
51 TADLERIKE VAGYEKVLQ REPDPNALQ GLEVARLQMG DLGAIAPME
101 KIKRIPERE GLKOLEIAIK LQYKNPOLPP PONPLAPTEP NNOGGTGGT
151 PTPPOP

IIA_SEQUENCE 1.0
PI:S74723 - hypothetical protein sl10939 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
A:Accession: S74723
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Mura, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

A:Reference number: S74322; MUID:97061201
A:Accession: S74723
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-125 <K&N>
A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017607; PID:g1651948
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S74723 Length: 125 February 11, 2000 15:52 Type: P Check: 8425 ..

1 MELIDHYLN LVSTIRFVLE SVSYLCILG LVKTLQWVA GDRRTMTPK
51 FENSTRLOFG LMLALAEFO LGADILSTTI APTLESISKL GLIAVIRTL
101 NYFLGKELES EIRMEIDQER QENGR

IIA_SEQUENCE 1.0
PI:S75160 - hypothetical protein sl11635 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
A:Accession: S75160
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Mura, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S75160
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-215 <K&N>
A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PID:d1017607; PID:g1652150
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S75160 Length: 215 February 11, 2000 15:52 Type: P Check: 4206 ..

1 MDVRFISLTK PELVIOGEP L SPEGLIAYCA RVSSPQENP NYTKLOFCI
51 REGHSITEM VDMLEITTI RAIAPQILRH RSFSFOEFL RYSCATEYEC
101 YEARRQDVKN RQNSLDDEDE STKRWFNQAO AAWWEKSHOL YEELAKGIA
151 KECARSILPL NVTYRLYMG SVRSWIRHFS VRCDQATQKE HREIALARK
201 IFMKHEPTVA AALEW

IIA_SEQUENCE 1.0
PI:S77220 - hypothetical protein slr1442 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
A:Accession: S77220
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Mura, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S77220

Mon Feb 14 08:07:32 2000

pir.cat

Page 140

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1285 <KAN>
A:Cross-references: EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018287;
PID:g1652634
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

S77220 Length: 285 February 11, 2000 15:52 Type: P Check: 3780

1 MLTVHRDARD YQIFLCSEFL SLGILARDWS LSPVIAVYL ATALTYOSTL
51 TINOQGNFSR WPSLDTWOSL WRHOGKRSAL ITGLGCLLL RANHLSTLIL
101 AASLAISAKR LLOAWGKHMFE NPANFGIIAA LTLTGDAWVS PGOKGTLML
151 GAVFLAAGGL VLKVGKRWDT SIMFLAVYAG LPLANAMUG WPEVSEFHL
201 ENGLSLVFL FMLTDRSIP NARSSRLIWA TAIATFSLIL QYHYFLPTAL
251 FVALFCLSPV TVLDROFPA PRFQWQKKN PVPSP

!!AA_SEQUENCE 1.0
P1:S77428 - hypothetical protein ser2060 - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998
C:Accession: S77428
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Natsu, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.
A:Reference number: S74322; MUID:97061201
A:Accession: S77428
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 179 <KAN>
A:Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PID:d1018264; PID:g1652610
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG

S77428 Length: 79 February 11, 2000 15:52 Type: P Check: 4795
1 MVLGNNLP TKIGAKLTMD SGRIVAVTG AIIILAVAY LVVYIIDFR
51 GENVPAPMET MGLLTHITAI ANGLSLPFS

!!AA_SEQUENCE 1.0
P1:B41838 - Vans - *Enterococcus faecium*
C:Species: *Enterococcus faecium*
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: B41838
J:Arthur, M.; Molinas, C.; Courvalin, P.
J. Bacteriol. 174, 2582-2591, 1992
A:Title: The Vans-Vanb two-component regulatory system controls synthesis of deipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147.
A:Reference number: A41838; MUID:92210502
A:Contents: BM4147; PIP816
A:Accession: B41838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1384 <AKT>
A:Note: sequence extracted from NCBI backbone (NCBI:94048, NCBI:94052)

B41838 Length: 384 February 11, 2000 15:52 Type: P Check: 3600

1 LVTKLNKNK DYSKLERLY MYIAVAVVA IYFVLYISM IRGLGCMIT
51 SIENNYDIN HDAKRLYOY SRRNIDIFI YVAIYISILI LCRWMSKFA
101 KYDEINTGI DVLQNEQDQ IELSAEMDVW ECKLTKART LERREDQAKL
151 AEOKNDVVA YLADIKRPL TSLIGYLSL DEAPMPYDQ KAYVHTLTD
201 KAYRLQID EFFELTRNL QITTLKTHI DLYIKLQMT DEEYPOLSAH
251 GKQAVIHAE DLTVSGDPK LARVENNIK NAAVSEDNS IDITAGLSG
301 DVVSIEFKNT GSIPKDLAA IFEKFRLEN ARSSDTGAG LGLIAKETII
351 VQGGQIYAE SNDYITFRV ELPAEDLVD KRRS

!!AA_SEQUENCE 1.0
P1:F69937 - conserved hypothetical protein ypj6 - *Bacillus subtilis*
C:Species: *Bacillus subtilis*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C:Accession: F69937
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Buisson, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Conerton, J.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoelt, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Follmer, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, G.; Galleon, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Goldlight, E.C.; Grandi, G.; Gutseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kraetz-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koeltter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.
S:Manuel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogilvay, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwilk, S.; Prescott, A.M.; Presseau, E.; Pujic, P.; Punelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.U.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Totsu, V.; Uchiyama, S.; Yamakoshi, M.; Yammer, F.; Vassarotti, A.; Viali, A.; Wandut, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: F69937
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1224 <KON>
A:Cross-references: GB:299115; GB:AL009126; NID:92634478; PID:e1183692; PID:92634665
A:Experimental source: strain 168
C:Genetics: ypj6
A:Gene: ypj6

F69937 Length: 224 February 11, 2000 15:52 Type: P Check: 3316
1 MYNADYVAFG ASDDVEIGM GGTAKFVKQ EKKVACDIT EALSSNGTV
51 SLREKAAEA ARIIGADKRI QLTIPDKGL MSDQAIRSV TVIRICPKKA
101 VPAFYKKDRH PDGNAALV EALIFSAGH KYKDEKSLPA HKYSKYVYM
151 INGHQPDFV IDISDTIAK KSLNAYKSO FTSPKDSVST PLNGYIEIV

201 EAREKTYGKE AGVEYAEVSE PNGC

!!AA_SEQUENCE 1.0
PI:539739 - efflux protein homolog ywf - Bacillus subtilis
N:Alternate names: protein lpa 84d
C:Species: Bacillus subtilis
C:Date: 07-Oct-1994 #sequence_revision 26-May-1995 #text_change 17-Mar-1999
A:Accession: S39739; GI:670052
R:Glasner, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzalez, M.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino, L.; Moszer, I.; Presecan, E.; Santana, M.; Schneider, E.; Schweizer, J.; Vertes, A.; Rapoport, G.; Danchin, A.
M1: Microbiol. 10, 371-384, 1993
A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees
A:Reference numbers: S39655; M01D:95020537
A:Accession: S39739; GI:670052
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <KUN>
A:Cross-references: EMBL:X7124; NID:9413923; PID:9414008
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1993
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Boesjes, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, C.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brouillet, S.; Burschler, C.V.; Cantwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.; Conerton, L.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Dusterhoedt, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Garsin, D.; Hahn, E.
Nature 390, 249-256, 1997
A:Authors: Rouger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Galleron, N.; Guin, S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, A.; Guiseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koettler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landinois, S.
A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Muehl, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tempira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, P.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; M01D:98044033
A:Accession: G70055
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <KUN>
A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PID:el186270; PID:92636306
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywf
C:Keywords: transmembrane protein
S39739 Length: 394 February 11, 2000 15:52 Type: P Check: 3644

1 KMOLEPNSKY LLYGALSFM GDYCVLPALL ILSTYDYDY VTSQVYVRS
51 IMWVOPFLG VLVDRIDRIK IMWTDING IIFLGLTFLP KSEYPLIFLA

101 LLEITYGSGV FENPARLAW SLESIDKSI NULFAATTI SIYGAAGG
151 LELGGSVEL AVAFNGVYL VSAFISRIK LOPVPOSEN IKENAFOSKE
201 GKREIKINSF VLNAMFTMT MALMGVYS YEPVSRFLG DGEIGNFLI
251 FQIGFGFTG AALYKXWGN NNRGITYTV LSIYSLALL FTPIPAVSI
301 AALFFIME YGEVLAKVY OENANOOG RITSVAESI GICISGSMF
351 INLSAPVIM GLIYVYGL FHTLVVKS FLERDNTQ KQVF

!!AA_SEQUENCE 1.0
PI:CA4365 - flagellar assembly protein fliH - Bacillus subtilis
N:Alternate names: fliH protein homolog
C:Species: Bacillus subtilis
C:Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 24-Sep-1998
A:Accession: CA4365; D69624; S14496
R:Albertini, A.M.; Caramori, F.; Crabb, W.D.; Scoffone, F.; Galizzi, A.
J. Bacteriol. 173, 3573-3576, 1991
A:Title: The fliH locus of Bacillus subtilis is part of a large operon coding for flagellar structures, motility functions, and an Arpase-like polypeptide.
A:Reference number: A42365; M01D:91258343
A:Accession: CA4365
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-208 <ALB>
A:Cross-references: EMBL:X56049; NID:939904; PID:939906
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Boesjes, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, C.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brouillet, S.; Burschler, C.V.; Cantwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.; Conerton, L.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Dusterhoedt, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Garsin, D.; Hahn, E.
Nature 390, 249-256, 1997
A:Authors: Rouger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Galleron, N.; Guin, S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, A.; Guiseppl, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karmali, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koettler, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landinois, S.
A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Muehl, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pulic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Tempira, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Vannier, P.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzengger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; M01D:98044033
A:Accession: D69624
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-208 <KUN>
A:Cross-references: GB:299112; GB:AL009126; NID:92633902; PID:el185214; PID:92633995
A:Experimental source: strain 168
C:Genetics:
A:Gene: fliH
C:Keywords: fliH
CA4365 Length: 208 February 11, 2000 15:52 Type: P Check: 1150

1 MARVEEADR ISEANSHE NRRQIDPOEK NDMAEKOKL IEAKAGEFE
 51 OGVALGRAEA MKOYALIQO ANITEMSRK AVEKDKEDAN EEIYELAVL
 101 AKRWQKSD DKEAFELLQO QVINEKVED DISIYDPY YETIPOKDE
 151 IOOLYKECR LGIYADEKAO KGTCTPTF GRVDSVDTQ LMOCKXLLT
 201 ALBNGAEE

!!AA-SEQUENCE 1.0
 PI:69820 - hypothetical protein yhay - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 05-Dec-1997 #sequence-revision 05-Dec-1997 #text-change 24-Sep-1998
 C/Accession: C69820

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Burschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, G.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, H.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwollik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Vassarotti, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viati, A.; Wanduit, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: C69820
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-83 <KUN>
 A:Cross-references: GB:299109; GB:AL009126; NID:92633260; PID:e1182983; PID:92633317
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yhay

C69820 Length: 83 February 11, 2000 15:52 Type: P Check: 5172 ..

1 MTSICLEOP FLAKKAMWS RLYWLVNDRI RSKOPFSLIS LFRFTKNHVS
 51 GPELTNDAL RLTLCLCLSP LYNGMIDRI HLK

!!AA-SEQUENCE 1.0
 PI:E69976 - hypothetical protein yrkG - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 05-Dec-1997 #sequence-revision 05-Dec-1997 #text-change 24-Sep-1998
 C/Accession: E69976

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Burschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, G.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, H.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwollik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Vassarotti, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viati, A.; Wanduit, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: E69976
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-76 <KUN>
 A:Cross-references: GB:299117; GB:AL009126; NID:92649466; PID:e1183681; PID:92650697
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yrkG

E69976 Length: 36 February 11, 2000 15:52 Type: P Check: 1724 ..

1 MYVAMPYKE YTKSKLTKNL YLFMVCWKAM ILVIGR

!!AA-SEQUENCE 1.0
 PI:S39656 - lipoteichoic acid biosynthesis protein dlte - Bacillus subtilis
 C/Species: Bacillus subtilis
 C/Date: 07-Oct-1994 #sequence-revision 26-May-1995 #text-change 17-Mar-1999
 C/Accession: S39656; G69616
 R:Glaser, P.; Kunst, F.; Arnand, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, M.; Lubochinsky, B.; Marcelino, L.; Moszer, I.; Pressecan, E.; Santana, M.; Schneider, E.; Schweizer, J.; Verites, A.; Rapoport, G.; Danchin, A.
 M1: Microbiol. 10, 371-384, 1993
 A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees
 A:Reference number: S39655; MUID:95020537
 A:Accession: S39656
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <GLA>
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993
 A:Cross-references: EMBL:X73124; NID:6413923; PID:6413925
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessières, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brigelli, S.C.; Bron, S.; Brulliet, S.; Burschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.R.; Codani, J.J.; Connetton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabre, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, G.; Galleron, N.; Ghm, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandt, H.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois, S.
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Mostl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetalle, D.; Porwollik, S.; Prescott, A.M.; Pressecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.; Ray, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Vassarotti, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viati, A.; Wanduit, R.; Wedler, E.; Wedler, H.; Wetzenecker, T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: S39656
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-76 <KUN>
 A:Cross-references: GB:299117; GB:AL009126; NID:92649466; PID:e1183681; PID:92650697
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yrkG

Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karameta, D.; Kashara, Y.; Klier-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinis, S.

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestli, D.; Nakai, S.; Nock, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Poll, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, T.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Seror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenberg, M.; Wambler, F.; Vassartelli, A.; Viari, A.; Wambler, R.; Wedler, E.; Wedler, H.; Weltergesser, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, Y.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033

A:Accession: G69618

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-252 <KUN>

A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PID:el166353; PID:92636389

A:Experimental source: strain 168

C:Genetics:

A:Gene: *dlte*

C:Superfamily: short-chain alcohol dehydrogenase homology

F:6-183/Domain: short-chain alcohol dehydrogenase homology <SADH>

S39656 Length: 252 February 11, 2000 15:52 Type: P Check: 2631

1 MKMTNTVLI TGSAGIGLE IAKRLLEGN EVIIGRSEA RLAEAKOQLP

51 NHTKQCDVA DSCREALYE WAKEXPNIN VLVNNAIGIK EIDFKKTEE

101 LTFDGEIEL NQAPVHLA LTFPHLMKOP EAAIVQVTS LARNPLAVP

151 VYCATAAALH SFTSLRHNV RDTSEVIEV APPNVDGLN QKSRKQGLT

201 YRGSSREYV QYFLDGKEG KQETINERVE GLRDATRADY DRLFQKNTQ

251 EN

!!AA_SEQUENCE 1.0

P1:572921 - hypothetical protein B2168_CL172 - *Mycobacterium leprae*

C:Species: *Mycobacterium leprae*

C>Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Sep-1997

C:Accession: S72921

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A:Description: *Mycobacterium leprae* cosmid B2168.

A:Reference number: S72580

A:Accession: S72921

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <SMI>

A:Cross-references: EMBL:U00018; NID:9467037; PID:9467073

S72921 Length: 86 February 11, 2000 15:52 Type: P Check: 2352

1 MTSINGPSAR DSAGKSVRDT GSEFGQPRP QFLTVAEVAL IMRVSKMTVY

51 RLVHNGELPA VRVGRSFRVH AKAVHDMLET SYFDVG

!!AA_SEQUENCE 1.0

P1:U06688 - coenzyme F420-dependent N5, N10-methylenetetrahydromethanopterin dehydrogenase homology - *Mycobacterium tuberculosis* (strain H37RV)

N:Alternate names: hypothetical protein RV2951c; hypothetical protein u0002p

C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Nov-1998

C:Accession: D70669; S73070

R:Colie, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998

Nature 393, 537-544, 1998

A:Authors: Sgares, K.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.

A:Reference number: A70500; MUID:98295987

A:Accession: D70669

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-381 <COL>

A:Cross-references: GB:Z83018; GB:AL123456; NID:93261671; PID:6283386; PID:91694883

A:Experimental source: strain H37RV

R:Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, September 1994

A:Description: *Mycobacterium tuberculosis* cosmid thc2.

A:Reference number: S73053

A:Accession: S73070

A:Molecule type: DNA

A:Residues: 1-49-194,197-381 <SMI>

A:Cross-references: EMBL:U00024; NID:9560506; PID:9560512

C:Genetics:

A:Gene: RV2951c

D70669 Length: 381 February 11, 2000 15:52 Type: P Check: 7905

1 MGLRFEFVD ALVHSRLPPT LPARSSMAA TWAGADSYW GDHNLALVPR

51 SIATSEVLGI AKFVPKIDA NYEPMTWLN LAFGLPSRLR LGVCTADAGR

101 RNPATVQAQA ATLHLITRGR ALDIGYGER EGNEPYGVEM TKPARPEEA

151 LATRALMNS NGELISRESP YFPLHNLFD LPPYRKWPE IWVAHCPRM

201 LRATGRYADA WPIPIVYRPS DYRALFAVR SAASDAGRPP MSITPAVRG

251 IITGRNDPV EEALSVYVK MTAIGVPEA WARRHVEHFM GADSGVQDI

301 IPOTMDKQTV LSYAAKVPAA LKKEVFSGT PDEVIDQAE WRDGLRIVV

351 LINGSLVNS LRKTVTAIPL HAKVIRGLIK L

!!AA_SEQUENCE 1.0

P1:C70828 - hypothetical protein RV0463 - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998

C:Accession: C70828

R:Colie, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998

Nature 393, 537-544, 1998

A:Authors: Sgares, K.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.

A:Reference number: A70500; MUID:98295987

A:Accession: C70828

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-97 <COL>

A:Cross-references: GB:AL021933; GB:AL123456; NID:g3261529; PID:e1252533;
PID:g2909539
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0463
C70828 Length: 97 February 11, 2000 15:52 Type: P Check: 2874 ..
1 MTRRASTDP QIMGAIGV VTEGYILMLAA ISVSDGLTIV SQMSRVLLL
51 SYLVAVCGAA GGLRLRSRGK LAWSAFAPSL PIPPVLTIVA VLADIYL
11AA_SEQUENCE 1.0
P1:D70510 - hypothetical protein RV0519c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: D70510
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krog, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: D70510
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <COL>
A:Cross-references: GB:297831; GB:AL123456; NID:g3261825; PID:e329668;
PID:g28266722
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0519c
D70510 Length: 300 February 11, 2000 15:52 Type: P Check: 2571 ..
1 MURRCAGNT DRGIMTPMA DLTRALLRW GAGAGAGAAG VMAFGALVDP
51 LEPQAPAPF EPTTAGSSLP TRISGSFISA ARGGIKTNVV ISMPGQSQ
101 LRPVIALGK DGNAMMLDL GVEQGLARLV KGRPAFAVV GVDGNTIWH
151 RSSGGDSGA MYLDELPLML TSMGMDTSRV GFLGSMGGY GALLGARLG
201 PARTAGICAI SPALFTSFTG STFGAFDSYD DYQHSYVLGL PALNSIPLRV
251 DCGTSDREYF ATRQFVNLH QPPAGSFSPG GHDSYVMEQ LPBELMAMS
11AA_SEQUENCE 1.0
P1:F70547 - hypothetical protein RV0547c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 12-Feb-1999
C:Accession: F70547
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krog, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: F70547
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-294 <COL>
A:Cross-references: GB:295558; GB:AL123456; NID:g3261781; PID:e316667;
PID:g2114009
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0547c
C:Superfamily: short-chain alcohol dehydrogenase homology
F:41-223/Domain: short-chain alcohol dehydrogenase homology <SADH>
F70547 Length: 294 February 11, 2000 15:52 Type: P Check: 3126 ..
1 MSKRLRWLN EQITLAGMRP PISPOLLNR PAMQPVLDLG KRILTGASS
51 GIGAAATKCP GLHRAVVAV ARKDLIDAV ADRITGDSGT AMSLPCLSD
101 MEALDALVD VEKRIIGIDI LINNAGSTR RLASLRLRW HDVERTVNL
151 YVAPRLIRG LAPGLERGD GHIINVAWG VISEASPLFS VYNSKRAALS
201 AVSRILTEW GSGGVHSTLL YPLVATPMI APTKAYDGLP ALTPAEAEW
251 MVTARTRPV RIAPRVAVAV NALDSIGPRV VNALMQRRNE QLNP
11AA_SEQUENCE 1.0
P1:C70708 - hypothetical protein RV0776c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: C70708
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krog, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: C70708
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <COL>
A:Cross-references: GB:280226; GB:AL123456; NID:g3261638; PID:e266566;
PID:g1550654
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0776c
C70708 Length: 259 February 11, 2000 15:52 Type: P Check: 4949 ..
1 MYEVGVDLIM AGRNPTGYAA VDAEDCLGVY GAARDASVL AALRPVYGD
51 CLVAFDAPLV VANRTGQRP EAALNDFRG FPAQAYPANT EKPEFADVP
101 AARLAROLAI DMDPLSSATR RAIEVYPHPA TVALFRLPRA LKTKAKDGS
151 VDLIKSELRL LMDGVGIAQ AGVAMOVAQ PDWVSLRQV TVAQKSDLR
201 AAEDPIDAVV CAYVALYQR RPADVTIYGD FTTGYIVTPS LPTDFRTAPD
251 AGRRARARR
11AA_SEQUENCE 1.0
P1:F70716 - hypothetical protein RV0953c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
A:Accession: G70716
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-294 <COL>
A:Cross-references: GB:295558; GB:AL123456; NID:g3261781; PID:e316667;
PID:g2114009
A:Experimental source: strain H37RV
A:Genetics:
A:Gene: RV0547c
C:Superfamily: short-chain alcohol dehydrogenase homology
F:41-223/Domain: short-chain alcohol dehydrogenase homology <SADH>
F70547 Length: 294 February 11, 2000 15:52 Type: P Check: 3126 ..
1 MSKRLRWLN EQITLAGMRP PISPOLLNR PAMQPVLDLG KRILTGASS
51 GIGAAATKCP GLHRAVVAV ARKDLIDAV ADRITGDSGT AMSLPCLSD
101 MEALDALVD VEKRIIGIDI LINNAGSTR RLASLRLRW HDVERTVNL
151 YVAPRLIRG LAPGLERGD GHIINVAWG VISEASPLFS VYNSKRAALS
201 AVSRILTEW GSGGVHSTLL YPLVATPMI APTKAYDGLP ALTPAEAEW
251 MVTARTRPV RIAPRVAVAV NALDSIGPRV VNALMQRRNE QLNP

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
 A: Reference number: A70500; MUID:98295987
 A: Accession: G70716
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-282 <COL>
 A: Cross-references: GB:Z79700; GB:AL123456; NID:93261628; PID:e264100;
 PID:g1524209
 A: Experimental source: strain H37Rv
 A: Genetics:
 A: Gene: RV0953c

G70716 Length: 282 February 11, 2000 15:52 Type: P Check: 883

1 MHYGLVLETS DRGITPAAA FLAESHGERT FYTPETHLP VKRQAHPTT
 51 GDASLPDDRY MRLDEPWSL GAASAVTSRI RLATAVALV EHDPTLAKS
 101 IATLDHLSHG RVSVGVFGM NDELVDHGV PGRRTMTLR EYLEMRLM
 151 TOEACDYDGE FVKFGPSWAM PKVPQHPVP LVGAAGTEKN EKWLRASDG
 201 WITPRVDVI DEPVKLLDI WAAAGRDGLP QIVALDVKVP PDKLARMAEL
 251 GTEVLEFGMP DRGADDAAY VERLAKLAC CV

!!AA_SEQUENCE 1.0
 P1:D70897 - hypothetical protein RV1102c - *Mycobacterium tuberculosis* (strain H37Rv)
 C: Species: *Mycobacterium tuberculosis*
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C: Accession: D70897
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
 A: Reference number: A70500; MUID:98295987
 A: Accession: D70897
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-103 <COL>
 A: Cross-references: GB:AL021997; GB:AL123456; NID:93260022; PID:e1251981;
 PID:g7896739
 A: Experimental source: strain H37Rv
 A: Genetics:
 A: Gene: RV1102c

D70897 Length: 103 February 11, 2000 15:52 Type: P Check: 9641

1 MRPHIAOLD KAPVILITR EVVAPHLTV TVAPITTVR GLATEVEVDA
 51 VNGINPSVY SCDNTQITPV CDLGROIGYL LASOEPALAE AIGNAFDLWM
 101 VVA
 !!AA_SEQUENCE 1.0

P1:F70958 - hypothetical protein RV1377c - *Mycobacterium tuberculosis* (strain H37Rv)
 C: Species: *Mycobacterium tuberculosis*
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C: Accession: F70958
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
 A: Reference number: A70500; MUID:98295987
 A: Accession: F70958
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-212 <COL>
 A: Cross-references: GB:Z81011; GB:AL123456; NID:93242274; PID:e275148;
 PID:g1621259
 A: Experimental source: strain H37Rv
 A: Genetics:
 A: Gene: RV1377c

F70958 Length: 212 February 11, 2000 15:52 Type: P Check: 4095

1 MPGIDFDALY RGEPSGEGLP PITTPPWDTK APRDNVIGMH TGGVHGVDL
 51 DIGCGLDGNA IYLRNNGYOV TGLDISPTAL TTAKRRASDA GVDVKEAVGD
 101 ATKRLGYTGA FDTVIDGMP HCLDDDGKRS YAASVHRATR PGATLLSCF
 151 SNAMPDEEM PRSTVSEQL RDLVGAGWD IESLEPAIVR RELDGEVEM
 201 AFMNVAQRGR GS

!!AA_SEQUENCE 1.0
 P1:C70763 - hypothetical protein RV1558 - *Mycobacterium tuberculosis* (strain H37Rv)
 C: Species: *Mycobacterium tuberculosis*
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C: Accession: C70763
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.
 A: Reference number: A70500; MUID:98295987
 A: Accession: C70763
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-148 <COL>
 A: Cross-references: GB:Z74020; GB:AL123456; NID:93261584; PID:e248339;
 PID:g1403484
 A: Experimental source: strain H37Rv
 A: Genetics:
 A: Gene: RV1558

C70763 Length: 148 February 11, 2000 15:52 Type: P Check: 497

1 MPLSGEAPS PLDMSPREQAD TYMSGCTEG TOLGKFPVLL LTVGARTGK
 51 LRKTPIMRVE HGGQVAIVAS LGGAPKNPVW YHVWVKNPVV ELDDGTVTGD

101 YDAREVFGDE KAIMWORA VMPDASYQT KTDROIPEV LFEVRAG

11AA_SEQUENCE 1.0
P1:C70541 - hypothetical protein Rv1577c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: C70541

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: C70541
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-170 <COL>

A:Cross-references: GB:295586; GB:AL123456; NID:93261785; PID:e317180;

PID:g2117260

A:Experimental source: strain H37RV

C:Genetics: Rv1577c

C70541 Length: 170 February 11, 2000 15:52 Type: P Check: 1907 ..

1 MAELRSEGR TVHGTVIPYN EATVRDPDG EPEMFAPGA FRSIAERGH

51 KILLVSHDA RTRYPVGRV ELREPHGLE GAFEIDPDP GDEALAVKA

101 GVVDSESVGF RPIRDRREGD VLVRVEALL EVSLGVPRV SGAQIAVRA

151 ESLTVASRST AEAWLSILDW

11AA_SEQUENCE 1.0
P1:A70639 - hypothetical protein Rv1957 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: A70639

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: A70639

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-181 <COL>

A:Cross-references: GB:284498; GB:AL123456; NID:93261701; PID:e293662;

PID:g1806205

A:Experimental source: strain H37RV

C:Genetics: Rv1957

A70639 Length: 181 February 11, 2000 15:52 Type: P Check: 9160 ..

1 MTRDADDDL DLORVGARLA ARAQIRDLRL LRTOAAVHRA EPRAGLIYD

51 LEFEAVADAD PATISATFVR ISCHIRIONO AADDVKEGD TKDETQVAT

101 ADEFFAALFD YHLOEGEDDP TEEELTAYAA TTGRFALYPR IREYVDLTG

151 RLALPELTLE ILSRPMFVSP GAQWPAETGRT P

11AA_SEQUENCE 1.0
P1:B70758 - hypothetical protein Rv1996 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: B70758

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: B70758

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-317 <COL>

A:Cross-references: GB:274025; GB:AL123456; NID:93261586; PID:e248804;

PID:g1403459

A:Experimental source: strain H37RV

C:Genetics: Rv1996

B70758 Length: 317 February 11, 2000 15:52 Type: P Check: 1836 ..

1 MSAQOINLGI VVVDSPSC HTAVENARD AQMRVALRV VQVVPVITA

51 PEGMAEYSR FOEAKREIV EHSYLAQAH QIVEQAKVA LEMSSGARA

101 QITGEVLHQ IYPTLANISR QVAMVVLGYR GQAVAGALL GSVSSTLVH

151 AHGPAVVIDE EPRPARPPA PVVVGIDGSP TSGLAELAF DEMSRGVDL

201 VALHANSDBG PLDPRLNWA PIERMLEDE QEKMLARLS QMODRPPVY

251 VHRVVCDBR APRLELAQT AOLVVGSHG RGFPGHNG SVSRVAVNSG

301 QAVVIARIP QDPAPVA

11AA_SEQUENCE 1.0
P1:A70941 - hypothetical protein Rv2018 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-Aug-1999
C:Accession: A70941

R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence

A:Reference number: A70500; MUID:98295987

A:Accession: A70941

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-239 <COL>

A:Cross-references: GB:AL021899; GB:AL123456; NID:93242282; PID:e1251997;

PID:g2896755

Mon Feb 14 08:07:32 2000

pir.cat

Page 147

A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2018
A70941 Length: 239 February 11, 2000 15:52 Type: P Check: 7719 ..

1 MAGDELELH EDVPLTIAE ASRYLVPRRA TLATWADGE RRPANAPVQ
51 GGPITLALPH PTGSHARLPF VGIAEAYVIN AFRRAGVPMQ RIRPSLMDLI
101 KNGVPHALAS QDLCTDGAEV LWFPAERSGE GSPDDLVVRG LIVPRSGQIV
151 FREIVHYIQ QISFADNLA SMIRLPQYGD ANVILDPRRG YGQPVFDGSG
201 VRYADVGLPL RAGATFQAVA DDYGVTPDOL RDAIDAIVA

11AA_SEQUENCE 1.0
P1:D70943 - hypothetical protein RV2037c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: D70943
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: D70943
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <COL>
A:Cross-references: GB:AL021899; GB:AL123456; NID:93242282; PID:e1252016; PID:92896774
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2037c
D70943 Length: 324 February 11, 2000 15:52 Type: P Check: 7125 ..

1 MALVSTARYD LVCEGGGVRG IGLVGNVDAI ADAGYRFPVY AGSSAGAIVA
51 SLVAALQTAG EPVTRLAEM RSIDYKPLD ENLIGHVPLI GGGSLILLSD
101 GYVRGAVLBD LLGGLLADLG VHTFGDLRTG EAPDFAMSL VTAASDLRR
151 RLVRIPMDLD SYGHPDDFS VARAVHASSA IPFVEPVRY RGATWVDSGL
201 LSNFPAVLFQ RTDAEPWMT FGIRLSRPF IPFTRPVQV VSLGIAIET
251 LVSNQDNAYI DDPCTVARI FVPAHVSPI DEDITAEORE ALYORGFQAG
301 OKFLANMNYA DCLADCGPFP TPSTL

11AA_SEQUENCE 1.0
P1:H70863 - hypothetical protein RV2449c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: H70863
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: H70863
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-419 <COL>
A:Cross-references: GB:AL021246; GB:AL123456; NID:93261507; PID:e1237574; PID:92791489
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2449c
H70863 Length: 419 February 11, 2000 15:52 Type: P Check: 4861 ..

1 MTATPREFDI VLXGATFVG KLAIEYLARA GGDARIALAG RSTQRYLAVR
51 EALGESAOYW PILTADASLP STLQAMARA QVYVTTGCV TRYGLPLVAA
101 CAAAGTDVAD LTGEPMERN SIDLYHKOA DIGARIYHC GFDSVPDLS
151 VVALYHAARE DGAGELTDIN CVRSFKGE SGGTISMLE VLASTANDPD
201 ARROLSDPYM LSPDRGAPE LQPODLPFR RGRRLAPELA GWTAGFLMA
251 PNTTRIVRS NALDWAYGR RFRYSRTWSY GSTVLAPVS VYGGVGNAM
301 FGLASVYRL LPRGIVKRVY PRGTGPSMA AREGYRIE TYTTTTCAR
351 VYARNQDD PGYKATSVLL GEGGLALAD RDLKLDWGV LTPAAMGDA
401 LRLRLPAGV SIQTRILAS

11AA_SEQUENCE 1.0
P1:E70572 - hypothetical protein RV2622 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: E70572
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: E70572
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-273 <COL>
A:Cross-references: GB:295387; GB:AL123456; NID:93261763; PID:e315941; PID:92104289
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV2622
E70572 Length: 273 February 11, 2000 15:52 Type: P Check: 8454 ..

1 MANKRGAG PPLSDRDD HMOGHWLAR LGRVLRPG VELTRILAR
51 AEVTDADVLE LAPGLGRTAA EILARNRSY VGEASDPNA NLVRYVLAR
101 GDVATDAD TGLSDASDV VIGEAALTMQ GNAKHITVA EAARVLRPG
151 RYALHELAV PDVAQVORT DLROSLARAL KYNARPLTVA EWSHLIAGH

```
201 LVEHVHTAS MALLOPRRI ADEGLGALR FAGNLIHRA ARRRVLMRH
251 TFRHRERLT AVAIVAKPH VDS

!!AA_SEQUENCE 1.0
p1:F70572 - hypothetical protein Rv2623 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: F70572
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: F70572
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-297 <COL>
A:Cross-references: GB:Z95387; GB:AL123456; NID:g3261763; PID:g316002; PID:g2104288
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2623

F70572 Length: 297 February 11, 2000 15:52 Type: P Check: 9524 ..

1 MSSGSSLSGI IVGIDSPAA QVAVEMARD AELRKIPITL VHAVSPEVAT
51 WLEVPILPPGV LRMQODGRH LIDDALKVE QASIRAGPPT VHSSEIVPAA
101 VPTLWMSKD AVLWVGCGLG SGRWPGRLG SVSSGLRHA HCPVVIHDE
151 DSVMPHPOA PVLGVGDSS ASELATIAIF DEASRRANDL VALHANSVDV
201 VSEMGIDWP ATGSMAGQVL AERLAGMER YPNVAITRV VROGPARQVL
251 ORSEEAQLVY VGRGRGGYA GMLVGSVGET VAQIARPPVI VARESLT

!!AA_SEQUENCE 1.0
p1:G70965 - hypothetical protein Rv2651c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: G70965
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: G70965
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-177 <COL>
A:Cross-references: GB:Z80225; GB:AL123456; NID:g3242265; PID:e266401; PID:g1550692
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2651c

G70965 Length: 177 February 11, 2000 15:52 Type: P Check: 4149 ..

1 MSSILFRTAE LRPEGERTVY GVIVPYGEVT TVRDLDEGR EMPAFGRFR
51 SIHERGKVA LVSHDARTY YPGRAVELR EEPHGLFAG ELANTPGDE
101 ALANVAKGV DAFSVGRPI RDRREGVIV RVEALBLEVS LTGPALGA
151 QIACVRAESL AVVRSLSADA RLALMDR

!!AA_SEQUENCE 1.0
p1:E70885 - hypothetical protein Rv2859c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: E70885
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: E70885
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-308 <COL>
A:Cross-references: GB:AL008883; GB:AL123456; NID:g3261490; PID:e1172951; PID:g2612804
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2859c

E70885 Length: 308 February 11, 2000 15:52 Type: P Check: 9649 ..

1 MDLSASRSDG GDLRPASPR LRSPVSDGCD PLRPSRPR SPVSDGGDPL
51 RPASPRLRSP LGASRPVGL TAYLEQVRTG VMDIRAGIYP ADYREGITMA
101 GCAVALIPQ PYDPESVGCY LDCIHALVIT GGYDDPRAVY GQEPHPATDH
151 PRPRDAMEF ALLRGALRG MPVLGICRGY QVYNALGCT LHOHLPDILG
201 HSGHRAQNGV FTRLPVHTAS GTRLAELIGE SADVCYHHQ AIDOVGEGLY
251 VSAVDVGVV EALTELPDGF VLAOVWPER SDDDLRFKA LVDAASVYAG
301 RQSAEPR

!!AA_SEQUENCE 1.0
p1:F70699 - hypothetical protein Rv2953 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 05-Dec-1998
C:Accession: F70699; S73068; S73067
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
```

A:Reference number: A70500; MUID:98295987
 A:Accession: F70669
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-418 <COL>
 A:Cross-references: GB:283018; GB:AL123456; NID:93261671; PID:e283385;
 PID:g1694681
 A:Experimental source: strain H37RV
 R:Smith, D.R.; Rodison, K. Library, September 1994
 A:Description: Mycobacterium tuberculosis cosmid tb02
 A:Reference number: 573053
 A:Accession: 573058
 A:Molecule type: DNA
 A:Residues:
 1-171, 'RAAFIATLIGRLGNSNFRSNHGIQRR', 'GPSARHRYVHADHGPGRRTFTWCAAFSSASRT'
 <SMIL>
 A:Cross-references: EMBL:U00024; NID:9560506; PID:9560514
 A:Accession: 573067
 A:Molecule type: DNA
 A:Residues: 'MNCVPTPT', 172-418 <SMIT>
 A:Cross-references: EMBL:U00024; NID:9560506; PID:9560515
 A:Genetics:
 A:Gene: RV2953
 F70669 Length: 418 February 11, 2000 15:52 Type: P Check: 5693 ..
 1 MSPAREEPDI VLYGATGFSG KLTAEHLAHS GSTARIALAG RSSERLGVR
 51 MMLGPAAMW PLILADASOP LILEMAARA QVVLTVGPY TRYGLPLVAA
 101 CAAAGIDYAD LTGELMFCRN SIDLYHKOAA DTGARILILAC GPDSPISDIN
 151 VGLYRASYE DGTGELCOTD LVLRSPQRM VSGGSVATYS EAMRTASSDP
 201 EARLYTDPY TLTTRGAEF ELGAQDPFLR RPRGLAPEL AGFWTGEFVQ
 251 APENTRYVR SNAIDEMAYG RFRYSSTMS LKSNAAPI LAAVGTIVAG
 301 TIGLKNKTFD RLPRLYERV TPKRGTPSR KTOERGHYTF EYTTTTTGA
 351 RYRATFANV DAYKSTAVIL AQSGLALAD RDRLAELGV LTPAALMGDA
 401 LLARLPAGV VMGTRLS
 1:AA-SEQUENCE 1.0
 P1:F70651 - hypothetical protein RV3073c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C:Accession: B70651
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: B70651
 A:Molecule type: DNA
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: B70651
 A:Molecule type: DNA
 A:Residues: 1-118 <COL>
 A:Cross-references: GB:283866; GB:AL123456; NID:93261691; PID:e290998;
 PID:g1781140
 A:Experimental source: strain H37RV
 A:Genetics:
 A:Gene: RV3073c

B70651 Length: 118 February 11, 2000 15:52 Type: P Check: 3649 ..
 1 MVRTRVRA RYEDIDPD GORLYDRIW PHGIRKDDOR VGWCKDVAP
 51 SKELREYVHH OPERDEFAS RYOEHLDSA ALAEELKLTG RSVVTPYAT
 101 RHVARSHAV LAQLNGR
 1:AA-SEQUENCE 1.0
 P1:D70645 - hypothetical protein RV3131 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C:Accession: D70645
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: D70645
 A:Molecule type: DNA
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: D70645
 A:Molecule type: DNA
 A:Residues: 1-332 <COL>
 A:Cross-references: GB:283867; GB:AL123456; NID:93261695; PID:e290954;
 PID:g1781236
 A:Experimental source: strain H37RV
 A:Genetics:
 A:Gene: RV3131
 D70645 Length: 332 February 11, 2000 15:52 Type: P Check: 2556 ..
 1 MNHFPDAET VRTVTLAVR APSIHNTQPM RMWVCPSTLE LFSRPMQMR
 51 STDPGRRLI LSCGVALLHC VALASLGMQ AKVNFPDPK DRCHLITIG
 101 QPLVPDQADY ALAAPIPRR TDRAYSCWP VGGDIALMA ARAARGVML
 151 RQVSALDRMK AIYAQAVLDH VTDEYLREL TWSGRYGSV AGVARNPEP
 201 SDSPAPIGR LFRAGPLSOP SDVLPADGA ATALGTEND DRLARLRAGE
 251 AASIVLTAT AMGLACPTT EPLEIATRD AVRAVFPAG GYPMMLRVG
 301 WAPINADLP PTPRELQSV VEMPELIRQ RC
 1:AA-SEQUENCE 1.0
 P1:F70951 - hypothetical protein RV3200c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C:Accession: F70951
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagers, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70951
 A:Molecule type: DNA
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70951
 A:Molecule type: DNA

A:Residues: 1-355 <COL>
A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PID:e1248791;
PID:g2827610
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3200C
F70951 Length: 355 February 11, 2000 15:52 Type: P Check: 8567 ..

1 MAGSWRLRG LNEKLTAPG VALVGLRIP QRASPARVI SRRVAVVA
51 LLTFAGIYV DRDGLDAG DRLTFLDLY YAAVTLSTG YGDIPISBF
101 ARAINIEYIT PLRIAFILIL VGTLEVLTE TSROAYKTOR WSRVNMHTV
151 VIGYTKGKT AVAMVSDLE VGEIYVVDI DSGYLERAAA AGLYVHGDA
201 TKSVDLRLAG TQHASIIVA TSRDCTAVLY TLTRRLIAPK AKIVASIRA
251 ENQHLLRQSG ADIYVVSSET AGRLLGIATF TPSVEMIED LLTPEAGLAV
301 AEREVEQAEV GGSFRLHDI VLGVVRDGL LRIAGPEVDA IEASDRLLYI
351 ROYGR

!!AA_SEQUENCE 1.0
P1:C70561 - hypothetical protein RV3626c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: C70561
R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
A:Reference number: A70500; MUID:98295987
A:Accession: C70561
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-350 <COL>
A:Cross-references: GB:295436; GB:AL123456; NID:g3261770; PID:e316560;
PID:g2105068
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3626c
C70561 Length: 350 February 11, 2000 15:52 Type: P Check: 48 ..

1 MTGASELTG NTVDWERFAS VGBRLARPAP PSTIYTRQV IDELIYAARK
51 ABPPVADVTG LIADGVVPPA RYVDRPAMIR SAASMRAMT HGSAPRGFL
101 TGRITGAOTG AVLAFAVSGI LQYDPFGAA GEGCLLVYP NYIAVEROLR
151 VEPSDFRLMV CLHEVTHRVQ FTANPLISG NSQALNLTF EPVDDIGRYV
201 SRLADIFSR GHGTDSEVN PGGILGIYRA VQSPQKAL DQLYLGTIL
251 BGHAHVADA VGPWVPSVA TIRRFDDRR HHKRPPIQRL VYALLGFDK
301 LQYTRGKAF VDHVDRAGM KLFNTIWGSP ETLPLPAEIE NPOKIDRYL

!!AA_SEQUENCE 1.0
P1:H70850 - hypothetical protein RV3912 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: H70850
R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
A:Reference number: A70500; MUID:98295987
A:Accession: H70850
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-254 <COL>
A:Cross-references: GB:AL021426; GB:AL123456; NID:g3261511; PID:e1245970;
PID:g2808697
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3912
H70850 Length: 254 February 11, 2000 15:52 Type: P Check: 8984 ..

1 MSADDPDK HSADAPPLT VELLADLOAG LLDATAARI RSRVSDPQA
51 QQLRALNRY RRDVAAGAD PANGPARPA VDSISAALR SARPNSSPA
101 AHARPHVHR VKMTAGAGL CAVATAIGV AVNDAPRPAP SATTQOHT
151 VSRPAPYPL SRPQYLDLH HTDYGPPG PLGDSRRTS CLSGLGTPAS
201 TPVLAGPID IDARPAVLV IPADTPDKLA VFAVAPHCSA ADTGLASTV
251 VPRA

!!AA_SEQUENCE 1.0
P1:H70865 - probable cutinase precursor - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: A70565
R:CoLe, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
A:Reference number: A70500; MUID:98295987
A:Accession: A70565
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-226 <COL>
A:Cross-references: GB:295390; GB:AL123456; NID:g3261766; PID:e316057;
PID:g2104375
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3452
A70565 Length: 226 February 11, 2000 15:52 Type: P Check: 5027 ..

1 MIPRPHSG RWRACARRL TSLVAAFAA ATLITPALA PRASACPPA
51 EVFARQTGE PGLGRVGOA FVSLROQTN KSIGTYGVN PANGDELAA
101 DGANDASDH OOMASACRAT RLVLGYSOG AAVIDIYTA PLPLGFTOP

```

151 LPPAADHIA AIALFQNSG RAGGLMSALT PQFGSKITNL CNNGPICSD
201 GNRMRALIGY VPCMTNOAR FVASRI

!!AA_SEQUENCE 1.0
P1:D70696 - probable dtdp-glucose 4 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Dec-1998
C:Accession: D70696
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: D70696
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-326 <COL>
A:Cross-references: GB:280343; GB:AL123456; NID:g3261648; PID:e267402;
PID:g1532867
A:Experimental source: strain H37RV
A:Gene: epib
A:Superfamily: UDPglucose 4-epimerase homology
F:3-31/Domain: UDPglucose 4-epimerase homology <UDP>
D70696 Length: 326 February 11, 2000 15:52 Type: P Check: 8388 ..

1 MEIIVTGAG FQSHLITSL LANKHWTVL DKSSNAYRN MQGRSDRA
51 AFISGSVTDG QTDRAVRDH HVEFLAAHV NPDQSLGDE SFLETNWGT
101 YRVLEAVRRY RNEIIVSTC EYGDGHNK EGERLDNAE LKPSPIGAS
151 KAADRILCYL YFRSYGLDVT TVRPNTIFV RQKAGFGAL IPRLVROGIN
201 GEGLTIFGAG SATRDYLYVS DIVAGYNLVL RPTFLRQAI NPAQKQTRV
251 RDIYEVADK FGARIEHRDA RGEVORFPA DISLANSIG QPOVELMDGI
301 DRYIMAKDQ PQYPEDDGE SGGSYVL

!!AA_SEQUENCE 1.0
P1:B70986 - probable isomerase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: B70986
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: B70986
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-203 <COL>
A:Cross-references: GB:295890; GB:AL123456; NID:g3242245; PID:e318986;
PID:g2131010

1 MTSRYRAPP IERYVILNDR GDATGVADKA TVHTGDTPLH LAESYVFDL
51 HDQILLTRRA AIKRTWPANW TNSCCGHPD GSLPGALIR RLAEALDLP
101 DRVDLIPGE RIRAMADGT VENICEYVR VQYDQPRN SDEVDALRWL
151 SWEQFVRDVT AGVIAPVSW CRSOLGYLTK LGRCPARQWV ADDCRLPKAA
201 HGN

!!AA_SEQUENCE 1.0
P1:H70760 - probable lipoprotein signal peptidase - Mycobacterium tuberculosis
(strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: H70760
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.
A:Reference number: A70500; MUID:98295987
A:Accession: H70760
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-202 <COL>
A:Cross-references: GB:274020; GB:AL123456; NID:g3261584; PID:e248552;
PID:g1403503
A:Experimental source: strain H37RV
A:Gene: lspa
H70760 Length: 202 February 11, 2000 15:52 Type: P Check: 2319 ..

1 MPDEPTGSAD PLTSTEDAG AGEPNAPAP RLRLMLLSVA VVYLTLDIV
51 KYVAVQLLPP GQPSIIGDT VTWLVNRSG AAFSNATGT WYLLIATGV
101 VVGIFWNGRR LVSPWALGL GMILGAMGN LYDRFRFAG PLRHWVDLF
151 SVGMWPYFNV ADPSVVGAI LVLITISIFG DEDTVGRHHA DGDIVGRKA
201 DG

!!AA_SEQUENCE 1.0
P1:B70763 - probable membrane protein - Mycobacterium tuberculosis (strain
H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
C:Accession: B70763
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.;
Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.;
Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagsels, K.;
Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail,
M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.;
Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.

```

A:Reference number: A70500; MUID:98295987
 A:Accession: B70763
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-397 <COL>
 A:Cross-references: GB:274020; GB:AL123456; NID:93261584; PID:e248540;
 PID:91403485
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: mmpL6

B70763 Length: 397 February 11, 2000 15:52 Type: P Check: 4418 ..
 1 MCGISVTVGLV KRGWVRVVE DTIDGIDQLG EQLASTVTVL DKLAIQPOL
 51 VALIPDEIAS QQINRELALA NYATMSGIYA QTRALLENNA AMGAFDPAK
 101 NDSFYLPE AFDPDPFQNG LKPLSADGK AARMTISHEG DPATPEGISH
 151 IDAIKQAAHE AVKGTPMAGA GIYAGTAAT FPDIDGATY DLTAGTIAL
 201 SLILIMMII TRSIVAAVVI VGIYALSGLA SFGLSVLMWQ HLGIDLYMI
 251 VLALAVILL AVGSDYNILL ISRKEEIGA GLNTGIIRAM AGTGVVYTA
 301 GLVEAATMMS FVPSDLRYLG QIGTTIGLGL LEDLLVYRAF MPESIAVLLG
 351 RWFVWQQRVA PRPASRMALRP YGPRPVVREL LLREGNDPR TOVATIR

!!AA-SEQUENCE 1.0
 P1:A70614 - Probable mma2 protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 17-Jul-1998
 C:Accession: A70614
 R:Coloe, S.T.; Broesh, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Besham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltham, D.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Hornsby, T.; Jagers, K.; Kirogi, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70614
 A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-287 <COL>
 A:Cross-references: GB:292772; GB:AL123456; NID:93261722; PID:e306575;
 PID:91877382
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: mmaA2

A70614 Length: 287 February 11, 2000 15:52 Type: P Check: 5098 ..
 1 MVNDLTPHPE DYQAHYDLSQ DFFRLPDPPT QTYSCAHER EDMTEBAQI
 51 AKIDIALGKL GLQPGMTLLD ICGGCGKATIR RAIQOYDVNV VGLTLKNOA
 101 AHYQSPDEK DTPRDRRYVL AGWEQNEPV DRIVISAGE HFHSHHADF
 151 FARAHKILPP DGVLLHRTIT GLTRQQNVDH GLPTLMLAR ELKFIATEIF
 201 PGQOPTIRM VEDOSAKTGF TLTRROSLP HYARTLDIMA EALOEHKSEA
 251 IAOSEYVEE RYKMYLTGCA KLFERYIDV NOFTLAK
 !!AA-SEQUENCE 1.0
 P1:S1338 - Hypothetical protein (insertion sequence IS112) - Streptomyces albus
 C:Species: Streptomyces albus

C:Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 30-Jun-1998
 C:Accession: S1338
 R:Rodicio, M.R.; Alvarez, M.A.; Chater, K.F.
 M.Ol. Gen. Genet. 225, 142-147, 1991
 A:Title: Isolation and genetic structure of IS112, an insertion sequence responsible for the inactivation of the SalI restriction-modification system of Streptomyces albus G.
 A:Reference number: S1338; MUID:91155926
 A:Accession: S1338
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <ROD>
 A:Cross-references: EMBL:X56644; NID:946594; PID:9581565
 A:Experimental source: insertion sequence IS112
 C:Genetics:
 A:Note: the authors translated the initiation codon GTG for residue 1 as Val
 A:Mobile element: insertion sequence IS112
 A:start codon: GTG

S1338 Length: 256 February 11, 2000 15:52 Type: P Check: 4079 ..
 1 MAGVTASEP SWIAPFSGLS PROFGKLVTV LRRGADAVR KGRPWSLPLE
 51 DRALLVAIW RTNLTMRQLA PLEGVSKSAA DRIDHLGPM LALQPKRFA
 101 KDTVLIVDGT LVPTRDHTIA ERSKNRYST NHQVVIDADT RLVVVVGHPIL
 151 AGNRNDCKAM EESGAKAAG KTLITADGY PGTGLVIPHR RERQAGSLPD
 201 WEEHNKSKH QVRAVEHVE ARMTWKILR DCRILKGQVH HAMGIAMNH
 251 NIALTG

!!AA-SEQUENCE 1.0
 P1:PC4110 - transcription regulator homolog - Streptomyces aureofaciens (fragment)
 N:Alternate names: hypothetical 137 protein
 C:Species: Streptomyces aureofaciens
 C:Date: 10-Jan-1996 #sequence_revision 08-Feb-1996 #text_change 09-Sep-1997
 C:Accession: PC4110
 R:Kotmanec, J.; Lempelova, A.; Farkasovsky, M.; Homeroova, D.
 Gene 165, 77-80, 1995
 A:Title: Cloning, sequencing and expression in Escherichia coli of a streptomyces aureofaciens gene encoding glyceraldhyde-3-phosphate dehydrogenase.
 A:Reference number: J04373; MUID:96084956
 A:Accession: PC4110
 A:Molecule type: DNA
 A:Residues: 1-137 <KOR>
 A:Cross-references: GB:U21191; NID:9706951; PID:9706952
 C:Comment: This protein has a helix-turn-helix DNA-binding domain in the carboxyl-terminus.
 C:Keywords: transcription

PC4110 Length: 137 February 11, 2000 15:52 Type: P Check: 2351 ..
 1 NHARLVAA PYRSGQAOY VPRSVPEVHG ERFAIREMA LRLGLEPIL
 51 ELARNAVS ARTFSRPAE DTGYTPDMWY MRARVDARE LERSRSVE
 101 QIADVGIST GSNRLHPOR ILGTPSEYR RIFTRGE
 !!AA-SEQUENCE 1.0
 P1:PN0644 - Hypothetical protein 66 - Streptomyces coelicolor (fragment)
 C:Species: Streptomyces coelicolor
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 07-May-1999
 C:Accession: PN0644
 R:May, J.T.; L.V.; Fisher, S.H.
 Gene 130, 145-150, 1993
 A:Title: The Streptomyces coelicolor glr gene encodes a protein similar to other bacterial response regulators.
 A:Reference number: JN0831; MUID:93345814
 A:Accession: PN0644

Mon Feb 14 08:07:32 2000

plr.cat

Page 153

```
A:Molecule type: DNA
A:Residues: 1-66 <MRA>
A:Cross-references: GB:L03213
C:Genetics:
A:Start codon: GTG

PN0644 Length: 66 February 11, 2000 15:52 Type: P Check: 9954

1 MAKVTRDDVA RIAGTSTAIV SYVINNGPPR VAPATREYVL AAKKEGYRP
51 DRYAQMASR RTDLIG

11AA_SEQUENCE 1.0
P1:S4423 - strp protein - Streptomyces glaucescens
C:Species: Streptomyces glaucescens
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Dec-1998
C:Accession: S4423
R:Mayer, G.; Piepersberg, W.
Submitted to the EMBL Data Library, April 1994
A:Reference number: S44224
A:Accession: S4423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-358 <MAX>
A:Cross-references: EMBL:X78974; NID:9475235; PID:9581674
C:Genetics:
A:Start codon: GTG
A:Superfamily: UDPglucose 4-epimerase homology
F:3-310/Domain: UDPglucose 4-epimerase homology <UDP>

S4423 Length: 358 February 11, 2000 15:52 Type: P Check: 6765

1 MRLITGOG YLGTWAPVL TAAGQVYGL DSGLFASRLV GALVDVYVG
51 LALDRDVTY DTLRGYDAV HLAALSNDEL GSDPELTHA VNHASTRLA
101 RLAKAGVTR FAYASTGSY GAQSGDAPVD EDAPLKYVP YAVSKYVED
151 DLVELADSGF APFLRNATA FGSPRLSD IYLNILGVA VLGRVTVLS
201 DGTAMPPLVH AEDTARAVLA GLAAPATYR ARAFNIGTE NNRVAEIAE
251 AAAAAYPGSV VEISGETGND PRSYRVDFTR ARTELGHRAI WIIPGAAQL
301 AREYRARGLT RAAREHDFTR LAVLERRRE GSLDGLDVR PARVRDGDH
351 RTATAEAG

11AA_SEQUENCE 1.0
P1:S17674 - serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) precursor -
Streptomyces sp. (strain K15)
N:Alternate names: D-alanyl-D-alanine carboxypeptidase; DD-peptidase;
penicillin-binding protein
C:Species: Streptomyces sp.
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Sep-1998
C:Accession: S17674
R:Palomagne-Messia, P.; Englebert, S.; Leyh-Bouille, M.; Nguyen-Disteche, M.;
Ducem, C.; Houbas, S.; Dideberg, O.; van Beumen, J.; Ghuyssen, J.M.
Bioscience 7, 229-230, 1991
A:Title: Amino acid sequence of the penicillin-binding protein / DD-peptidase
of Streptomyces K15. Predicted secondary structures of the low-M(r)
penicillin-binding proteins of class A.
A:Reference number: S17674; MUID:92028812
A:Accession: S17674
A:Molecule type: DNA
A:Residues: 1-291 <PAL>
A:Cross-references: EMBL:X59965; NID:947150; PID:9581684
C:Genetics:
A:Start codon: GTG
A:Keywords: hydrolase; serine carboxypeptidase
F:1-29/Domain: signal sequence #status predicted <SIC>
F:30-291/Product: D-alanyl-D-alanine carboxypeptidase #status predicted <MAT>
```

```
S17674 Length: 291 February 11, 2000 15:52 Type: P Check: 4817

1 MLRLRAATV ITTGALLAAG TLGATPATAV TKPTIAVGS YAMNGTGT
51 LTKKADIRR STGSTKIMT AKVVALQSNL NLDKATIQK AYSDDVANK
101 PSQAHLYVD KVTYROLLG LMLPSGCCDA YALADKSGS SQAAVRKSF
151 IGMNTAATN LGLNTHFDS FDSIGGANV STPHRTKIA SSAMKNSYFR
201 TVVKTKAYTA KTVTKGSIR TMDTKNTNG LLSSYSAIG VTKGSGPEAK
251 YCLVFATRG KTVIAGTVLA STSPARESD ATKIMNGFA L

11AA_SEQUENCE 1.0
P1:S17717 - rRNA methylase - Streptomyces tenebrarius
N:Alternate names: gentamicin resistance protein
C:Species: Streptomyces tenebrarius
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Mar-1998
C:Accession: S17717
R:Holmes, D.J.; Cundliffe, E.
Mol. Gen. Genet. 229, 229-237, 1991
A:Title: Analysis of a ribosomal RNA methylase gene from Streptomyces
tenebrarius which confers resistance to gentamicin.
A:Reference number: S17717; MUID:92017656
A:Accession: S17717
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <HOL>
A:Cross-references: GB:S60108; NID:9237470; PID:9237471
C:Keywords: antibiotic resistance

S17717 Length: 280 February 11, 2000 15:52 Type: P Check: 5969

1 MPRAPGPGD PEDPRLAEVY DAVSSRRYQ SVAPETVRRL ATSLAVASRG
51 DLAEAVKRTK RGLHEIFGAY LSPSPKYDAL IROLRGAVDA ATTRPGGHPA
101 PRHYARLHP RALPIDEFY REVFARCAP ASYVDIACGW NPLAPMWP
151 SDAFTYHSD IDTRLMEFLD AALETLGVAN DAVRYDLIMG VGEVETDVL
201 LKTVPCIEA QGRGQWDLI DAIRSPLYVY SFPRKSIGOR SKGMENYSA
251 NEDAWLENRP HDVQLEFRN ELVYFRKNA

11AA_SEQUENCE 1.0
P1:C64251 - hypothetical protein homolog MG464 - Mycoplasma genitalium (SGC3)
C:Species: Mycoplasma genitalium
C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 10-Oct-1997
C:Accession: C64251
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.;
Fleischmann, R.D.; Bult, C.J.; Kellaway, A.R.; Sutton, G.; Kelley, J.M.;
Fittichman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.;
Nguyen, D.; Uiterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb,
J.F.; Dougherty, B.A.; Boff, K.F.; Hu, P.C.; Lueder, T.S.; Peterson, S.N.;
Smith, R.O.; Hutchison III, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MUID:96026346
A:Accession: C64251
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-385 <TRGR>
A:Cross-references: GB:U39733; GB:L43967; NID:91046177; PID:91046184; TIGR:MG464
A:Experimental source: strain G-37
C:Genetics:
A:Genetic code: SGC3

C64251 Length: 385 February 11, 2000 15:52 Type: P Check: 2836

1 MPKIAQINK EITTFNPFM SAAVNEKNN WKNFKESAI FIKYKVFIF
```


51 IFLITVLGNG CTOTLAQPMW GTNOVLGSL EIGYKFGTGG DYRYDLISNN
 101 FGPFITSDY TLAYGPFYGM FWPMAQIVL PIMYATRPVL GSGVELGFNM
 151 ILSLIVLLL VRLITIVITL NSTLAEKMN EVOGKLAELN AKYGAIDLQ
 201 SKNRQLEIM SLYKKNIKS SAAVQVEVT LPFLIIRI VTLRPIKAI
 251 ILENFMDLSK VPLTEIFSNF TTGWPFIIF LVIIVPVQFL SOKLPQVMS
 301 KRNENAKAHS OKSIEDLNKT KKMOLIEFYV FAIATAFSAA GVGWYFELNA
 351 LFTLOSITL HVEIVKRRK RKONYSKLD ILERE

!!AA_SEQUENCE 1.0
 P1:164230 - sensory rhodopsin II transducer homolog - Mycoplasma genitalium
 C:Species: Mycoplasma genitalium
 C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Oct-1997
 C:Accession: 164230

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Duster, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; MUID:96026346
 A:Accession: 164230
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-248 <TIGR>
 A:Cross-references: GB:U3707; GB:U43967; NID:91045970; PID:91045975; TIGR:MG280
 A:Experimental source: strain G-37
 A:Genetic code: SGC3

164230 Length: 248 February 11, 2000 15:52 Type: P Check: 1048

1 MUYVFLGFG GITATYFRST AANLYVOARN SIDSSFNASK AFANALANSA
 51 NQEKSSITN NLDQVKKLE OSLOKVDYK KMLESONNNG NISOEKIREL
 101 DAIKKDLIENS KTOLDNFKNM LDKNGZASS PSVKKATAD GYLSAVSEPS
 151 TQAOISIVSY EKIKNNIPSS EGFNNYDYVT MITIVAASGG MLAILITITV
 201 FSEITSKRKG LIRFSRFIST EQLADHVNDI LDRYPDLEE VITALDHD

!!AA_SEQUENCE 1.0
 P1:S73488 - hypothetical protein K05_orf385 - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 17-Jul-1998
 C:Accession: S73488
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.
 A:Reference number: S73327; MUID:97105885
 A:Accession: S73488
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-385 <HIM>
 A:Cross-references: EMBL:AE000017; GB:U00089; NID:91673812; PID:91673823
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Genetic code: SGC3

S73488 Length: 385 February 11, 2000 15:52 Type: P Check: 5866

1 MPJLNKKK ELKTFNPFW SAAVNEKNG LKNLKKAMGI IFRYLKVAIF
 51 IFLITVLGNG CTOTLAQPMW GTNOVLGSL EIGNFGTGG DYRYDLOSNN
 101 VGPFITSDY TLAYGPFYGM FWPASQIVL PIMYATRPVL GSGVELGFNM
 151 ILSLIVLLL VRLITIVITL NSTLAEKMN EVOGKLAELN AKYGAIDLQ
 201 SKNRQLEIM SLYKKNIKS SASVQVEVT LPFLIIRI VTLRPIKAI
 251 ILENFMDLSK VPLTEIFSNF TSTGWFIF LVIIVPVQFI SOKLPQVMS
 301 KRNENAKAHS OKSIEDLNKT KKMOLIEFYV FAVITAFSAA GVGWYFELNA
 351 LFTLOSITL HVEIVKRRK RKNYSKLD ILERE

!!AA_SEQUENCE 1.0
 P1:A36149 - spiralin - Spiroplasma melliferum (SGC3) (fragment)
 C:Species: Spiroplasma melliferum
 C:Date: 30-Nov-1990 #sequence_revision 13-Jan-1993 #text_change 31-Oct-1997
 C:Accession: A36149
 R:Chevalier, C.; Salllard, C.; Bove, J.M.
 J. Bacteriol. 172, 6090-6097, 1990
 A:Title: Spiralin of Spiroplasma citri and Spiroplasma melliferum: amino acid sequences and putative organization in the cell membrane.
 A:Reference number: A36149; MUID:91008990
 A:Accession: A36149
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-237 <CHE>
 A:Cross-references: GB:M59366; GB:M33991; NID:9152891; PID:9152892
 C:Genetics:
 A:Genetic code: SGC3
 C:Keywords: Transmembrane protein

A36149 Length: 237 February 11, 2000 15:52 Type: P Check: 5138

1 SILAVFVSA VGTTSVACN KTESNLSRV KTIAPATVA ASTPRANVRP
 51 EIKTALNAV LKAVGVYKT AFAADQFEV YKNSKGALF TIDLNGAVE
 101 VYLQITPARD KTVVIGERY IKVILPKHGE VTKVDIVV VREQTVGKA
 151 STPKAVKDE LNAVNTYATL AKAVLDATON IAPNAGSDP EITNNGAEGD
 201 YEAKREVEVT VKAKNDSANI SQGFKRAKV TATAPTE

!!AA_SEQUENCE 1.0
 P1:H69019 - conserved hypothetical protein MTH1148 - Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
 C:Accession: H69019
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delongher, C.; Lee, H.; Dubois, J.; Aldridge, T.; Bashirzadeh, R.; Blakey, D.; Cook, R.; Gilbert, R.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, L.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; Modougall, S.; Shimer, G.; Goyal, A.; Petrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.
 A:Reference number: A69000; MUID:98037514
 A:Accession: H69019
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-157 <MTH>
 A:Cross-references: GB:AE000884; GB:AE000666; NID:92622242; PID:92622253
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1148

H69019 Length: 157 February 11, 2000 15:52 Type: P Check: 3145 ..

1 MIRATACEF THGFAREH AVSNGYPGY SMSVSDVVL VAGLFIPTLS

51 GRSILFEF PEPSTATINDI KYTEEDER VALMARSVR ELTSDLDIG

101 TTACIGGCI AVSENREY INSDVADLR FSGAEIILR QRSGRCLAE

151 LLESFLE

!!AA_SEQUENCE 1.0

PI:D69049 - conserved hypothetical protein MTH1371 - Methanobacterium

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998

C:Accession: D69049

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;

Alredede, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison,

D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.;

Vicatore, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.;

Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.;

Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.;

Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta

H: functional analysis and comparative genomics.

A:Reference number: A69000; MUID:98037514

A:Accession: D69049

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-355 <MTH>

A:Cross-references: GB:AE000900; GB:AE000666; NID:g2622478; PID:g2622480

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1371

D69049 Length: 355 February 11, 2000 15:52 Type: P Check: 1234 ..

1 MNLKVAWE LRSTLSRKF LFIPLQAV LILIFMFG FIDVLESGBA

51 LTPSLTGRAE LGVIDPSGLI QGQNPDLVD IREGDAGSV LYVDNFTGIP

101 LNALIYLDYS DPRSYVRDE VEAVERASS LISEFVKSP GPDPEVREIT

151 RGEPLPQVL TRVWSTILF LPEVLEGLV VDSIVGKER KTGELIAMP

201 VRSEIILGK CLSVITVAL OIGVMNILL ASGFHMVNP L GAYFTVYASS

251 APVAGLIALI SVYARNYREA GIGITLAYII SAATLIPAL AYMASSGSV

301 SPMTLTITKI SGOAGLPTS FLQPSQILYS ASYSMALQSG CSGGMWTSV

351 PDQVY

!!AA_SEQUENCE 1.0

PI:D69065 - conserved hypothetical protein MTH1489 - Methanobacterium

C:Species: Methanobacterium thermoautotrophicum

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998

C:Accession: D69065

R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.;

Alredede, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison,

D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.;

Vicatore, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.;

Bush, D.; Saefer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.;

Goyal, A.; Pietrokowski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.;

Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta

H: functional analysis and comparative genomics.

A:Reference number: A69000; MUID:98037514

A:Accession: D69065

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-249 <MTH>

A:Cross-references: GB:AE000909; GB:AE000666; NID:g2622596; PID:g2622605

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH1489

D69065 Length: 249 February 11, 2000 15:52 Type: P Check: 9991 ..

1 MRLGFSTLAL FMEPLENIQ KAEDDGFELY ELICGEPYWP RLLQDGLSL

51 EYFESFDLEY LIHAPITDNL PASMNGIRE EGRGMVETI ELMSRGATV

101 VTHRPGVVR REDRRSAL QFALFELGEC VEAADLSIK FSVENMPGRF

151 SYLCNPAEH ERFVKGCSY ATVDIGHANT TGRLODFLEI KRANHYISD

201 NNGKROHLP LGEYTVDLKL LGSIERGVIE LNSYGVIRS RLILEVVR

!!AA_SEQUENCE 1.0

PI:G64362 - 2-Isopropylmalate synthase (EC 4.1.3.12) - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: G64362

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton,

G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage,

A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;

Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, D.; Nguyen, D.; Ullrich, T.R.;

Geoghegan, N.S.M.; Feldman, J.F.; Fuhmann, J.L.; Nguyen, D.; Ullrich, T.R.;

Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,

K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;

Moese, C.R.; Venter, J.C.

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus

jannaschii.

A:Reference number: A64300; MUID:96337999

A:Accession: G64362

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-406 <BUL>

A:Cross-references: GB:U67500; GB:L77117; NID:g1591202; PID:g1591206;

TIGR:MJ0503; PID:g1510577

C:Genetics:

A:Map position: FOR444536-445756

C:Keywords: carbon-carbon lyase; oxo-acid-lyase

G64362 Length: 406 February 11, 2000 15:52 Type: P Check: 6767 ..

1 MTKYLVKMD FLFENSKAV CPYPRKLDK DIYIDTILR DGBOTPEVCF

51 TKECKLEIAR KIDELGLKQI EAGFPYVSR EADYKTIAN EGNADILAL

101 CRALKDKIDK AIECDYDGI TFIATSPHL KTKENKSID EILEMGYBAV

151 EYAKHGLEY AFSADATRT PIEDLIKVRK AAEAGADRV HIADTGCAT

201 PQSNEFTICK LKENIKRAHI GVCHNDGFG AVINSTYGLI GKAAVSTTV

251 NGIGERAGNA ALEELIMALT VLYVDVGLN LEVLEPLCKM VEESGKIMP

301 KNRPIVEVY FKHESGIHVD AVIENPLTYE PLPEKIGLK RNILGKHSRG

351 CRAVAYKLK MGIDYDREML CEIVKVKET REEGKEITDE VKKEIVEVL

401 KKRKNK

!!AA_SEQUENCE 1.0

PI:G64456 - hypothetical protein homolog MJ1251 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997

C:Accession: G64456

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273: 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.
 A:Reference number: A64300; MUID:96337999
 A:Accession: D64456
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-167 <BUL>
 A:Cross-references: GB:U67565; GB:L77117; NID:g1591874; PID:g1591884;
 TIGR:MJ1251; PID:g1511255
 C:Genetics:
 A:Map position: FOR1194471-1194974

C64456 Length: 167 February 11, 2000 15:52 Type: P Check: 7004 ..

1 MITATACCF THANIGLTIH KAAAGYEDFE FKYLEBEDL KLMKNVRVYS
 51 AMFVSIIGV EKLDIRLPE PDFNYKYAKA YSEKDEVA KLMAGGLAKK
 101 LNVNLSIGST AGVGRGALICI LTDNNRYLFT SDVYANLITF ENIKEROKNG
 151 IEKGKRFLE ILKREYF

11AA_SEQUENCE 1.0
 P1:D64423 - hypothetical protein MJ0988 - *Methanococcus jannaschii*
 C:Species: *Methanococcus jannaschii*
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
 C:Accession: D64423
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273: 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.
 A:Reference number: A64300; MUID:96337999
 A:Accession: D64423
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-329 <BUL>
 A:Cross-references: GB:U67541; GB:L77117; NID:g1591641; PID:g1591650;
 TIGR:MJ0988; PID:g1511021
 C:Genetics:
 A:Map position: FOR918363-919352

D64423 Length: 329 February 11, 2000 15:52 Type: P Check: 9981 ..

1 MELLEYLRKD EVELFCHHNA DDDAVGSCVA LKYLASQNP NGKFRISADS
 51 VSKLSRNILN EIGERVDIET YPKLPETVFI VDRASINOLK VNFDELKERE
 101 VILIDHKKKT DLADICKYYI IKEDYPTSE IIAEIFKELN IPPPNVRIA
 151 LLGGIVDTK HLLANSKTF ELISYLINDI SFQKILYLS QESDVKRFA
 201 HLKCSRMEI REPDKLRIAL SHVSSHEASC AKTIVSIGAD VAFVAVVRKK
 251 EKEIRVSNRC RKHVSKYVHL GNLMEKIKKE LGGSGGSHSE AGGLNAPVDK

301 SKSEKVIKE VLNLCYKRFV EBYKKAON

11AA_SEQUENCE 1.0
 P1:D64452 - hypothetical protein MJ1221 - *Methanococcus jannaschii*
 C:Species: *Methanococcus jannaschii*
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 10-Oct-1997
 C:Accession: D64452
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273: 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.
 A:Reference number: A64300; MUID:96337999
 A:Accession: D64452
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-299 <BUL>
 A:Cross-references: GB:U67563; GB:L77117; NID:g1591850; PID:g1591852;
 TIGR:MJ1221; PID:g1511223
 C:Genetics:
 A:Map position: FOR1164064-1164963
 A:Start codon: TTG

D64452 Length: 299 February 11, 2000 15:52 Type: P Check: 2920 ..

1 MTFGEKRRYM KIIPKFLN TVEILIKRNN AVSISIEPL KISIEGIIII
 51 TCNADARDAE KIYELKKLG LGEKGHGYT IMPANTIFSC REGGIASTSL
 101 SPEELTYKAK TWKTKITNVI IKVILASNG VIGLIEHNIP TLIGAMITAP
 151 LVDTVNGSAI GTVLGDELF IQGKRELIC SGIVYCAFI PSIFVSKEL
 201 VLQYLSETSI ILSAIVAIIA GISGGSINS GKEYEIIQVT IDVSLIIPAL
 251 LMGMALATMD LYLITITFIL LAIIVLIDV GYIGLKYKV GKINOKIKY

11AA_SEQUENCE 1.0
 P1:F69396 - acyl-CoA dehydrogenase, short chain-specific (acds) homolog - *Archaeoglobus fulgidus*
 C:Species: *Archaeoglobus fulgidus*
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
 C:Accession: F69396
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, G.; Ketchum, K.A.; Dodson, R.J.; Gwin, D.E.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyriakides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenry, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McConaughy, L.
 Nature 390: 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujili, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.
 A:Reference number: A69250; MUID:98049343
 A:Accession: F69396
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-437 <KLE>
 A:Cross-references: GB:AE001023; GB:AE000782; NID:g2689346; PID:g2649411;
 TIGR:AF1175

F69396 Length: 437 February 11, 2000 15:52 Type: P Check: 9110 ..

1 MWMKPYLID DELSEEDRL KEVHFRAEE VIRPASTELD RMPPEERKVP
 51 GSPFKWKWK IKKLGYHMY LPENKGGPGL TPLORILFE EIANGSLGFA
 101 TALGVNDMPF VSAAFGCTPE IYEDLVYPMI EDESGFHGC WGVTEPHGS
 151 DYLITLREEP ELVAKFGKN VYAEKDGDW VISGQSAWV SSAPVATHC
 201 LHAOLKGRS IAEGLFCIYP LDADGVKKR PADMLGMRD POGELFFDGV
 251 RIPEHYVVA PGFFYGVFD OLLCLTSCG GAFVGLAAR CEELLYHAR
 301 ORVGGVDIV KHKVKIKLT EMFEVEVETAR YYVKAETETI HKKILEQCTA
 351 DASPHARAA QIYAKIATE VAHDLQJFG AYGLIKFVI EKFFDAGCM
 401 LIEDGIVEVL SLEAEDVID NYEKERYDVN HYKAVLG

!!AA_SEQUENCE 1.0
 P1:G69440 - conserved hypothetical protein AF1528 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
 C:Accession: G69440
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
 A:Reference number: A69250; MUID:98049343
 A:Accession: G69440
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1136 <LRE>
 A:Cross-references: GB:AE000997; GB:AE000782; NID:g2689320; PID:g2649036; TIGR:AF1528

G69440 Length: 136 February 11, 2000 15:52 Type: P Check: 1599 ..

1 MDVNYALLQT LLMKRVPI MLVSLILSNA PFLSPQFSRV GEFINNIREF
 51 SKLSGRFAIA SFLIPIVGI SLSANFRAG KLSFRITACS AALNMLRGF
 101 RAALFLAPV AVSVFGEIG LFEVSLDIS RLAVFL

!!AA_SEQUENCE 1.0
 P1:A69470 - hypothetical protein AF1762 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
 C:Accession: A69470
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, J.H.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
 A:Reference number: A69250; MUID:98049343
 A:Accession: A69470
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A69470 Length: 61 February 11, 2000 15:52 Type: P Check: 6749 ..

1 NNPILKRDG KLAETILNRP EVHNFESWAT LTRGAEKVV TMLKRRRNS
 51 TEIKFLCRNA F

!!AA_SEQUENCE 1.0
 P1:D69509 - hypothetical protein AF2077 - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-Jun-1998
 C:Accession: D69509
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kervage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.
 A:Reference number: A69250; MUID:98049343
 A:Accession: D69509
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-270 <LRE>
 A:Cross-references: GB:AE000960; GB:AE000782; NID:g2689283; PID:g2648461; TIGR:AF2077

D69509 Length: 270 February 11, 2000 15:52 Type: P Check: 4730 ..

1 MKRLILIAA TLVLVSSG NFRXYGARG VAVNIADNNS SYIGFCEPEM
 51 TLYIANDST GVISVKNNG EDVLYFSTP NDILFISNPT FLYAGEERKL
 101 EAERGSQGE YIIPIDIEAF WDNGSAKIPA CSKVYVDPAY RMSKVLDSGN
 151 TVPLFTREV WKFRILVESD AGDNYTILDY IPAEFEVLST AASDGSFATH
 201 HPGAGRSCTT KIWDVYVGC TEPMDVTIAT KINPAGKQEF TSPGSYMLND
 251 GAERIGLGIY SNPIYKAVR

!!AA_SEQUENCE 1.0
 P1:T08352 - hypothetical protein H1337 - Halobacterium sp. (strain NRC-1)
 C:Species: Halobacterium sp.
 A:Variety: strain NRC-1
 C:Accession: T08352
 R:Ng, W.Y.; Cliffo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baslin, D.; Faust, J.; Hall, B.; Loretz, C.; Seto, J.; Slagel, J.; Hood, L.; DasSarma, S.
 Genome Res. 8, 1141-1144, 1998
 A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or mitochondrion?
 A:Reference number: Z16408; MUID:99063795
 A:Accession: T08352
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-148 <DAS>
 A:Cross-references: EMBL:AF016485; NID:g2822278; PID:g2822413
 A:Experimental source: strain NRC-1
 C:Genetics: plasmid pNRC100

Mon Feb 14 08:07:32 2000

pl.r.cat

Page 158

```
T08352 Length: 148 February 11, 2000 15:52 Type: P Check: 7002
1 MSNSPEETHD FDMFPGVD GEGCIRIARA DREYSGFGE LAPMLRITHE
51 QLTGLDAEG WIGVKVONS EXYVDHRLQP RIEVTEYRD HLMQALCAYC
101 DTHDVNCHVS SLEPNEDRSD QYMGVRSIN NTKITLTPLR DQFYKRE

11AA:SEQUENCE 1.0
P1:S73242 - hypothetical protein 38 - red alga (porphyra purpurea) chloroplast
C:Species: chloroplast Porphyra purpurea
C>Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 10-Sep-1997
C:Accession: S73242
R:Reith, M.; Munholland, J.
A:Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast
genome.
A:Reference number: S73108
A:Accession: S73242
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-291 <REI>
A:Cross-references: EMBL:U38604; NID:q1276652; PID:q1276787
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October
1995
C:Genetics:
A:Genome: chloroplast
C:Keywords: chloroplast

S73242 Length: 291 February 11, 2000 15:52 Type: P Check: 4681
1 MRFAPSKKE LKPIKEDTP KYVSYIIIE IEALWORLD QVWRPATIM
51 AGIOPPLMW VFGGLFCNA PVNLFITNS YNRELSGII VPSFGALN
101 SGLPMFDRE FGLNRLILA PLISRTSIIF SSATFMTCLS LIQVIFIVA
151 SLFPGNSPLS NSSTLIFALI VLLVTVGM LSLALSTLP GHIELLALIL
201 VVNLPLFESS TALAPLYFMP PWLOLIASIN PLSYAIKIR YIYSMTDMWF
251 TESVAKISWG DISLOIISL LFLDVIGAY IVSNILKARL N

11AA:SEQUENCE 1.0
P1:S20471 - class V zygote-specific protein - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Mar-1998
C:Accession: S20471
R:Matters, G.L.; Goodenough, U.W.
Mol. Gen. Genet. 232, 81-88, 1992
A:Title: A gene/pseudogene tandem duplication encodes a cysteine-rich protein
expressed during zygote development in Chlamydomonas reinhardtii.
A:Reference number: S20471; MVID:92204139
A:Accession: S20471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86 <MAT>
A:Cross-references: GB:S90874; NID:9247969; PID:9247970

S20471 Length: 86 February 11, 2000 15:52 Type: P Check: 4134
1 MKGFAKLIA LAILATFSA THAGPIYGI CQTGNALAY ACYAAAGTF
51 GYPSWGAAP ATVLACNAGL GTCMACVAA GUSPIP

11AA:SEQUENCE 1.0
P1:S25969 - hypothetical protein 69 - liverwort (Marchantia polymorpha)
mitochondrion
C:Species: Marchantia polymorpha
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 17-Mar-1999
C:Accession: S25969
R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takamura, M.; Nozato, N.;
Akashi, K.; Kanehara, T.; Ogura, Y.; Kohchi, T.; Ohgama, K.

J Mol. Biol. 223, 1-7, 1992
A:Title: Gene organization deduced from the complete sequence of liverwort
Marchantia polymorpha mitochondrial DNA. A primitive form of plant
mitochondrial genome.
A:Reference number: S25941; MVID:92114051
A:Accession: S25969
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-69 <ODA>
A:Cross-references: EMBL:M68929; NID:9786182; PID:9786194
A:Note: the nucleotide sequence was submitted to the EMBL Data Library,
February 1992
C:Genetics:
A:Genome: mitochondrion
C:Keywords: mitochondrion

S25969 Length: 69 February 11, 2000 15:52 Type: P Check: 7679
1 MASLFSFRT QFLTPATF IYPAHPSHF GFAFGQLH LFGPGGLPE
51 RENAHCHID NKRTPRLQ

11AA:SEQUENCE 1.0
P1:T01531 - hypothetical protein A_IG005110.4 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01531
R:Andrews, S.
A:Submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005110.
A:Reference number: Z14347
A:Accession: T01531
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-289 <AND>
A:Cross-references: EMBL:AF013293; NID:92252823; PID:92252827
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Insertions: 156/1
A:Note: A_IG005110.4

T01531 Length: 289 February 11, 2000 15:52 Type: P Check: 1750
1 MVAHAKSATI IPKERLKNRI VEHGRLAQR PTPLANIITY LWFPGFILS
51 IIRYFNPLP PERFARYTE MLAGHLITRG HRPSPSPGT LGNLYVYNHR
101 TALDEIYAI ALGRKICVT YSVRLSML SPIPAVALTR DRATDANMR
151 KLEKGLDVI CPEGTGCEE YLAFSLFA ELSDRIVPA MNCKQNGNG
201 TTVGAVFMD PYFFFNPRP SYEATFDRL PEDKTVNGGG KTPLEVANYV
251 QKVGAVLGF ECTEUTRKDK YLLLGNDGK VESINTRK

11AA:SEQUENCE 1.0
P1:T05405 - hypothetical protein F10M6.130 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05405
R:Bevan, M.; Weichselgartner, M.; Farlmann, B.; Granderath, K.; Dauner, D.;
Heizl, A.; Neumann, S.; Hohsels, J.; Nemes, H.W.; Mayer, K.F.X.; Schueller, C.
A:Submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15414
A:Accession: T05405
A:Molecule type: DNA
A:Residues: 1-222 <BEV>
A:Cross-references: EMBL:AL021811
A:Experimental source: cultivar Columbia; BAC clone F10M6
C:Genetics:
A:Map position: 4
A:Insertions: 39/1; 100/1; 131/3; 172/2
```

Mon Feb 14 08:07:32 2000

pir.cat

Page 159

A:Note: F10M6.130

T05405 Length: 222 February 11, 2000 15:52 Type: P Check: 4069

1 MTGLVSLFA SVERNREELT IWRVEODDE EHWICVCEET LICVAMELO
51 LGRGRSVTA GRCESSLMR SKKSEGTASF AVETVLMAGF AGHFWYRGA
101 TLERLALSTRA IDEKSKIRL LKREVAIEL KPSNPLPION DKAILGDVGI
151 PLHRQOYTT RAHRTIRRT ISPLMGLSHE LVRRFGSTIE IAIGGSGRVR
201 KGLPFKEDL MKNSNEDD LG

!!AA_SEQUENCE 1.0
P1:T00559 - hypothetical protein F12L6.15 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse ear cross)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 30-Apr-1999
C:Accession: T00559
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A:Reference number: 214168
A:Accession: T00559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-296 <ROU>
A:Cross-references: EMBL:AC004218; NID:93355463; PID:93355478
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Note: F12L6.15

T00559 Length: 296 February 11, 2000 15:52 Type: P Check: 9966

1 MEHRKDLIS CLPLELLIYI ISFLPESAR LPLVSTRFR SVNQAALIVA
51 HSHNGIEDI SHAVSRFINN FDEHDSKNT RLELHVDS TFWSTILAPH
101 NVMHSGFFS GSKKEESFC WRLEIDDQIR RRVDSGFLV KTLCLDSVNS
151 LTRHVVSNV LEFLSDSLK ICCKRRLSL TIDSPKLLH LSISGCPKLR
201 YLELISFKIK TPHYGSLPL IKIHENLT KAVPDVTCGP SYNNALDIG
251 PLSLTKNSQ SLTICRMFE VTKLVSKFEC SKIKRLIYV IFVNRK

!!AA_SEQUENCE 1.0
P1:T01617 - hypothetical protein F19F24.9 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 14-May-1999
C:Accession: T01617
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
submitted to the EMBL Data Library, April 1998
A:Description: Arabidopsis thaliana chromosome II BAC F19F24 genomic sequence.
A:Reference number: 214153
A:Accession: T01617
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-392 <ROU>
A:Cross-references: EMBL:AC003673; NID:93004543; PID:93004564
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Note: F19F24.9

T01617 Length: 392 February 11, 2000 15:52 Type: P Check: 1414

1 MKYIRNSLK RLFSFKRSF DSDSENSSP LAASIKVE GFQETDOFOR
51 PKMKCFQOE IYDATINGESS ENLVGRGGA EYKGLGN GEELAVKRIT
101 RGRDRERE KEFLMEIGTI GHVSHPNVLS LIGCCINDGL YLVFESSRG
151 SLASLHDLN QAPLEWETRY KLAIGTAKGL HYLKRCQQR IHRDIRSN
201 VLLNDFEPQ ISDFGLAKWL PSQMSHSLA PIEGTGHLA PEYTHGIYD
251 EKTIVFARCV FLELLSGKK PYDASHQSLH SWAKLIKDG EIEKLVDPRI
301 GEEDLQQLH RIAFASICI RSSSLCRPSM IEVLVLOGE DIEKRWKME
351 EEEVKEEFW GYEDLEDC EC DSSISLSPD SISNRSSSHR SR

!!AA_SEQUENCE 1.0
P1:T06062 - hypothetical protein F19H22.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T06062
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 215184
A:Accession: T06062
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <BEV>
A:Cross-references: EMBL:AL035679
A:Experimental source: cultivar Columbia; BAC clone F19H22
C:Genetics:
A:Map position: 4
A:Note: F19H22.120

T06062 Length: 169 February 11, 2000 15:52 Type: P Check: 59

1 MYSKPIGPF AFRRHQHLL LRHLGIFSK EBNFLQVDE GGMIVPLLT
51 IIVYTKILM ACCCECCQ MEDDAYSEWL RANGVNAS QNHGTSNDAM
101 GYFLGEVWF YQADSDFELS LSVGDYVIR EVVSVWMEG ECKNGMFT
151 YVIERDRV FAKVIEVF

!!AA_SEQUENCE 1.0
P1:T04272 - hypothetical protein F20B18.280 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
C:Accession: T04272
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hohnselt, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 215263
A:Accession: T04272
A:Molecule type: DNA
A:Residues: 1-349 <BEV>
A:Cross-references: EMBL:AL049483
A:Experimental source: cultivar Columbia; BAC clone F20B18
C:Genetics:
A:Map position: 4
A:Note: F20B18.280

T04272 Length: 349 February 11, 2000 15:52 Type: P Check: 5780

1 KQILIGANEK ENFREKDLV GNRVQGAFO GLVELSHDG RRDVLVNL
51 GQPESTRRL RSYRSFAHH DLLKGLSQT ILPTQKSD NOTERKSDS
101 EEEREVSSDA AEKESNLSPL ILRLSRQ PVSEKHDDIV DSDSASAG
151 VLEEDGTCT TTPYGRKRC TEHGRRLSR VSPDIHICE VPVRECEET

Mon Feb 14 08:07:32 2000

pir.cat

Page 160

```
201 ENIGGVLLPD MIRCRSKPVV RRRCEBHG MRVNAEFFLL NPTERKAVN
251 EDKRPETST GANQEGSGL CEATTKNGLP CTRSAPEGK RCMQHKDRL
301 NHGSENVQS ATRAVICGF KLYNGSVCK SPVGRKRCE EHKGRITS

!!A_SEQUENCE 1.0
P1:T05639 - hypothetical protein F20D10.240 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C/Accession: T05639
R:Bevan, M.; Medler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.;
Mayer, K.F.X.; Scheller, C.
Submitted to the Protein Sequence Database, February 1999
A:Reference number: 215420
A:Accession: T05639
A:Molecule type: DNA
A:Residues: 1-445 <BEV>
A:Cross-references: EMBL:AL035538
A:Experimental source: cultivar Columbia; BAC clone F20D10
C:Genetics:
A:Map position: 4
A:Introns: 88/3; 148/2; 234/3; 283/1; 399/3
A:Note: F20D10.240

T05639 Length: 445 February 11, 2000 15:52 Type: P Check: 5981

1 MFTAAASSV GRMRTAFSL RDEISTTPEP PVPLLEDDL FQQSHSLISA
51 VSHLPHLHET SDCLFLLDLV SKADGPDWIP VSRHTCOLIH DVCARLLFOL
101 NSSMPFLIH SFASVLEFLL RQPMSPSPS AAFVSRIPEV IQCFETLRL
151 APMHEPHLH VKFLVRYVPL LHQDLVLSYG FSNODSEPTL IVEKRLPOON
201 RLMSMALAF DMFGRAFSLS ESLFPTVSO CTLEVLARKVA DVLASKGOLV
251 EDREMRKMP LVLMRLQFTF FELGISRLVA LLASLMEFC FELTGPQLS
301 VSDVYHNDK LNVKISPLIS GVSNAKNTF YRPPHRRKD DLNTRPVSS
351 SWRLSAHDS GSSDVISDS DFSDSDGSP DSYFAQSKV RLAIVCIQD
401 LQADSKSFT TQWVLEPTS DVLRPKWIKIN LQHLSPYS IFLSP

!!A_SEQUENCE 1.0
P1:T01144 - hypothetical protein F26B6.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C/Accession: T01144
R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
Venter, J.C.
Submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.
A:Reference number: 214198
A:Accession: T01144
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-272 <ROU>
A:Cross-references: EMBL:AC003040; NID:93242700; PID:93242730
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 131/3; 177/3
A:Note: F26B6.20

T01144 Length: 272 February 11, 2000 15:52 Type: P Check: 6323

1 MLVRKAMEN KNORFVLIH GVCHGAMTWK KVKTOLEVAG HCVTVNDLAA
51 SGIMTKVVE IOTLNDYCKP LLEFLSIGS DDGKVIYVAH SNGGISALA

201 ADSFACTIA IYELIEMPD TINPAYVE KILRSIPOE WLDTCVNG
151 KPDPFLQYTL LGFKEMAKM YONSPVODLE VYKTLVRENP LVTNLAGTR
201 SFSEGVGSV TRIIVYCRD LVEVEDYORW MISNPPREV MEIKADHMP
251 MFSKPOEVA LLEIANKXC KN

!!A_SEQUENCE 1.0
P1:T01145 - hypothetical protein F26B6.21 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C/Accession: T01145
R:Roundley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon,
Venter, J.C.
Submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome II BAC F26B6 genomic sequence.
A:Reference number: 214198
A:Accession: T01145
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <ROU>
A:Cross-references: EMBL:AC003040; NID:93242700; PID:93242731
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 124/3; 167/3
A:Note: F26B6.21

T01145 Length: 260 February 11, 2000 15:52 Type: P Check: 9711

1 MDKNOKKFTV LVHGICHGAM CWYKVAOLE AAGHSTYAVD LAAGVNNTS
51 LDEIOTLNDY CKPLLEFLS LGSDDDKVIL VASMGISIA SLADIPFSK
101 VAAIVFAAF MPDISNPAY VFQKLVKDVY QEVMQDVGK KPDPLEFAL
151 FGPERMAYTL YNLSPLQDFE LAKMSVRVSP FMNNLAGTI SFSDRGVSV
201 TRIIVYCGD VAVPYDORG MINDFPVKEV LEIKADHMP MFSKPOLCA
251 LLEIADKXA

!!A_SEQUENCE 1.0
P1:T08549 - hypothetical protein F27B13.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C/Accession: T08549
R:Bevan, M.; Zimmermann, W.; Gruenelsen, A.; Wambutt, R.; Bancroft, I.; Mewes,
H.W.; Mayer, K.F.X.; Scheller, C.
Submitted to the Protein Sequence Database, May 1999
A:Reference number: 216442
A:Accession: T08549
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <BEV>
A:Cross-references: EMBL:AL050352
A:Experimental source: cultivar Columbia; BAC clone F27B13
C:Genetics:
A:Map position: 4
A:Introns: 89/3
A:Note: F27B13.120

T08549 Length: 404 February 11, 2000 15:52 Type: P Check: 4270

1 MICEPAYOYO QHGOXDHMM TMAMDLSTS PPSSPLSPSL SPKSSYNNN
51 EERLEVVNL SGMALQSLPN PELNLANICK LDISNNHIK IPESITALL
101 NLIADISHN QIKALPSIG CISKILINV SGNFIVSLQ TIQNCRFTS
```

151 TSGGNSYAM MIMINSVF FCGFGRSL ELMANENELI RLPDNGIEL
 201 TNRKLCVNS NKILSPATI TGLTSTRVLD ARUNCMLIP EDLENILNE
 251 LINVSONFOY LSLPSSIGL LNLLEFLDS YNKITVLPES ICGMRRLRL
 301 SAEGNPLVSP PIEVEOSLH AVREISQMA NGKLVNNAK KTWGRKRLV
 351 KYGTENGSR VWTREEREGL IMPERYPIDI LTRKFPVTC SPRRLSPRT
 401 YFSR

11AA_SEQUENCE 1.0
 P1:T05096 - hypothetical protein F28M20.70 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T05096
 R:Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hohnesfel,
 J.; Mewes, H.W.; Meyer, K.F.X.; Schueller, C.
 Submitted to the Protein Sequence Database, November 1998
 A:Reference number: Z15398
 A:Accession: T05096
 A:Molecule type: DNA
 A:Residues: 1-171 <BEV>
 A:Cross-references: EMBL:AL031004
 A:Experimental source: cultivar Columbia; BAC clone F28M20
 C:Genetics:
 A:Map position: 4
 A:Introns: 30/1
 A>Note: F28M20.70

T05096 Length: 171 February 11, 2000 15:52 Type: P Check: 9032 ..

1 MDRLEAFMY LLSLETINOS EVEAVALP SADASASRNI YDMAEKLTYQ
 51 SISAVTPGVK NLSSDOQLP VARYEALTD GRPNPDESY IILDARASKS
 101 GSIDNSYVK GPEBALYFMI GGGNIEFSS LQELSQOEM VNNITGATE
 151 ILTGTELVQ LGELGQRMGL R

11AA_SEQUENCE 1.0
 F1:T02589 - hypothetical protein T16B24.23 - Arabidopsis thaliana (fragment)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 23-Apr-1999
 C:Accession: T02589
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 Submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
 A:Reference number: Z14679
 A:Accession: T02589
 A:Status: translated from GB/EMBL/DBJ
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-381 <ROU>
 A:Cross-references: EMBL:AC004697; NID:93402671; PID:93402690
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A:Introns: 60/2, 169/1, 240/3, 248/3
 A>Note: T16B24.23

T02589 Length: 381 February 11, 2000 15:52 Type: P Check: 3673 ..

1 MNHVPSQSF YTESEDEDR KDYYEDDGS HSDSSDYVD ENQAHKPPSS
 51 YTTAMPQSYR QSIDLYSSVP SPGIGFLGN SMTRFSSFL SSGILRRHTP
 101 ESLPTVTKPL LEOADEQAL PKHRLSSGL LSPIPRRRS MRKDKSSMV
 151 SHEIPMSRNS SYGQAVLNGI NVLCGVGIS TPTAANGGM LGIMLIFYTG

201 LLSFTYGLL RYCLDSESDL ETPDIGOA FGTGRIEVS IVLXELTYAC
 251 CYEYILLED NLSSLYPNA LSIGFOLDA RHLFALLITL AVLPTWLRD
 301 LSVLSTISNG GVIVSVLV CLFWIGLVD VGHSKGTTL NULSTLPAIG
 351 LVGYCSGHA VFPNITSMA KSOYPAVL T

11AA_SEQUENCE 1.0
 P1:T02510 - hypothetical protein T19C21.16 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 14-May-1999
 C:Accession: T02510
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 Submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.
 A:Reference number: Z14676
 A:Accession: T02510
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-96 <ROU>
 A:Cross-references: EMBL:AC004683; NID:93395421; PID:93395437
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A:Introns: 17/3
 A>Note: T19C21.16

T02510 Length: 96 February 11, 2000 15:52 Type: P Check: 8930 ..

1 MSRYLRSHA WCLRQNVSN KSFLEKSELY GLGYRPIRY FIDIGPLPMW
 51 TRKEYNELIL GRTRVSPIL KLILFLVHR ICSDPRFOVK SSSVNA

11AA_SEQUENCE 1.0
 P1:T02122 - hypothetical protein T20B5.17 - Arabidopsis thaliana
 N:Alternate names: hypothetical protein T3K9.26
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 14-May-1999
 C:Accession: T00761; T02122
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 Submitted to the EMBL Data Library, November 1997
 A:Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.
 A:Reference number: Z14159
 A:Accession: T00761
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-248 <ROU>
 A:Cross-references: EMBL:AC002409; NID:92623294; PID:92623311
 A:Experimental source: cultivar Columbia
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 Submitted to the EMBL Data Library, February 1999
 A:Description: Arabidopsis thaliana chromosome II BAC T3K9 genomic sequence.
 A:Reference number: Z14570
 A:Accession: T02122
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-248 <ROU>
 A:Cross-references: EMBL:AC004261; NID:93402695; PID:93402720
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A>Note: T20B5.17; T3K9.26

T02122 Length: 248 February 11, 2000 15:52 Type: P Check: 8329 ..

1 MREDNWFML RWEELPSP ELPISQTLI TPLALAFQI GSPNHHGSK


```

51 RTTAIVHQKL QSSTPTPTPT PTPPPMMNMS DFGGSDTDL GSGSIGEPA
101 RTLKRPLVW TPQLHKRFVD AVGHGIGTKNA VPKTIQGLMS VEGTLREWVA
151 SHLOKRYLYL RRMGGGNGNG ITGSHVIVSD SADRFLFASS PVPAHFLSPD
201 YLMPDEHSHY MGRHVTQON QVFNLRXED SEYHGSGMKM LKLPFAGN

!!AA_SEQUENCE 1.0
P1:T02348 - hypothetical protein T26B15.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 23-Apr-1999
C:Accession: T02348
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
Submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
A:Reference number: Z14678
A:Accession: T02348
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-217 <R0U>
A:Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298537
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 25/1
A>Note: T26B15.5

T02548 Length: 217 February 11, 2000 15:52 Type: P Check: 5743 ..

1 MKRGIRHLCE NGFTGYSLH HHYHEHHVDF FEEBLIVYT HTPSVIRMT
51 HSIRFVSRIR LSHPLVGLG VQWTPRGSD PPDIQLCVG TRCLIIOLSH
101 CKYVPDLRS FLEDQITIFV GWNNSQDKR LERFHOLDI WRLVHIRYL
151 HPLLSSFE TIVKYLGHE GVTKDKELCM SNMGARSISH DQIVQASHDV
201 YVCKKLGKVE RLKMKGA

!!AA_SEQUENCE 1.0
P1:T05859 - hypothetical protein T29A15.30 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999
C:Accession: T05859
R:Bevan, M.; Van der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Hohnsels, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
Submitted to the Protein Sequence Database, March 1999
A:Reference number: Z15455
A:Accession: T05859
A:Molecule type: DNA
A:Residues: 1-208 <BEV>
A:Cross-references: EMBL:AL035602
A:Experimental source: cultivar Columbia; BAC clone T29A15
C:Genetics:
A:Map position: 4
A:Introns: 151/2
A>Note: T29A15.30

T05859 Length: 208 February 11, 2000 15:52 Type: P Check: 7427 ..

1 MAFSPRLSL SVPPDAFESW LRDSGYIELL DHRTSAAAA ASSASVSSS
51 AAATSAASD VVSTITGGEF ASILSRIVY SELLTINFS KLSADDFSGD
101 TTPWTGFIIG NCDYSFSS SQAARVHE NIKRPARNYA TLFTVFACA
151 LYQMPALVG LIGSLIMEL FKYSCKWKF DRHPSMRKLS IGIGCGEFP
201 DHEVPVFC

!!AA_SEQUENCE 1.0
P1:T01365 - hypothetical protein T29F13.9 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C:Accession: T01365
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
Submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence.
A:Reference number: Z14177
A:Accession: T01365
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-175 <R0U>
A:Cross-references: EMBL:AC003096; NID:g3132469; PID:g3132475
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Introns: 73/1
A>Note: T29F13.9

T01365 Length: 175 February 11, 2000 15:52 Type: P Check: 1733 ..

1 MGLVTKALKY LLLSLSLF IYTNVSSAS PMTPSSPAK MSRLVAVEG
51 MYCKSCSKYS GVDTLLEASP LOGATVKIAC NNTKRGVME TKTDKNGYFF
101 MIAPKKITTY AFTCRAMPY NPGPTATMT CTYPSKLNNG ITGAMLKPSK
151 TINIGHDYV LFSVPFAFE PACAL

!!AA_SEQUENCE 1.0
P1:T08903 - hypothetical protein T32A16.10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08903
R:Bevan, M.; Zimmermann, W.; Gruenelsen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schueller, C.
Submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16518
A:Accession: T08903
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <BEV>
A:Cross-references: EMBL:AL078468
A:Experimental source: cultivar Columbia; BAC clone T32A16
C:Genetics:
A:Map position: 4
A:Introns: 153/1; 325/1
A>Note: T32A16.10

T08903 Length: 364 February 11, 2000 15:52 Type: P Check: 2816 ..

1 MNTCTIVSEP KTHSSLASLK KLVLSGANFS AETESISFTN KSCITYIDVS
51 KTSLSKNSFL ETWFNLEHD LSTAFGDDS VGFVACVGEN LKNINVSDTQ
101 ITPGAVGNLA GHVPOLETLS MSQTFVDDL ILISTWPC IKALDLGNMS
151 TLGFIYLISP QEEKSKSLAA LQSLTSLTEL SLEHPTLGDK ALSGLSLTG
201 LTHSLTSTS LNDSTLHLS SLPLVSLGV RQCVLTISGL EKFPNNRLR
251 TLIDQGCWLL TKDDIAGLCK RYPIKVRHE HADSSLDON QLDPRSSTPQ
301 SFGVYARRNN QRESSVAV RSFLDQVRY NREELVALON SPLSLPWE
351 ELVSVPELLA DSVI

!!AA_SEQUENCE 1.0
P1:T06094 - hypothetical protein T5J17.50 - Arabidopsis thaliana
```

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999
 C:Accession: T06094
 R:Bevan, M.; Murphy, G.; Riddle, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.;
 Mayer, K.F.X.; Schueller, C.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: 215184
 A:Accession: T06094
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-178 <BEV>
 A:Cross-references: EMBL:AL035708
 A:Experimental source: cultivar Columbia; BAC clone T5J17
 C:Genetics:
 A:Map position: 4
 A:Note: T5J17.50

T06094 Length: 178 February 11, 2000 15:52 Type: P Check: 7672 ..

1 MGSRLGRVY HFANLPKIL MPKLTNIHE FALKTIPSAS KIEIKRVLES
 51 LVGFVDEKYN TLNMDGKKK RGGLIYAKAD YKAYVTLRS PLISRDLPF
 101 VKYIEDRKS KYGGSFVEE EDDKSHWLD RAEKREIIGY GKSGRRGGE
 151 RANSPTRGGA AAGTAKFPW SNMRFVCK

11AA_SEQUENCE 1.0
 P1:T00988 - hypothetical protein T9J22.22 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 14-May-1999
 C:Accession: T00988
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
 S.M.; Kaul, S.; Mason, T.M.; Kellavage, A.R.; Adams, M.D.; Somerville, C.R.;
 Venter, J.C.
 submitted to the EMBL Data Library, April 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T9J22 genomic sequence.
 A:Reference number: Z14153
 A:Accession: T00988
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-241 <ROU>
 A:Cross-references: EMBL:AC002505; NID:92739359; PID:92739380
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A:Introns: 159/1
 A:Note: T9J22.22

T00988 Length: 241 February 11, 2000 15:52 Type: P Check: 916 ..

1 MASLAPPL LSTPRKLTLS HLHTSISPF QISTORRPOK HLINICRSTP
 51 TPQQAASOR KRTYRKQYP GENIGITEEM RFVAMRLNV NGKIDLSFD
 101 KITEKEEEE EEDDDDDDE VKDETWRPSK EGFLKYLVD KLVETIERI
 151 VDESENVSYA YFRRTGLERC ESEKDLWL REODVLEP SNVGSYARY
 201 LEOAGESAP LFLSHFYSIY FSHIAGGYL VQVSEFFGL S

11AA_SEQUENCE 1.0
 P1:T08191 - probable translation elongation factor ts - Arabidopsis thaliana
 N:Alternate names: protein T22B4.100
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
 C:Accession: T08191
 R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer,
 K.F.X.; Schueller, C.
 submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16098
 A:Accession: T08191
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-415 <BEV>
 A:Cross-references: EMBL:AL049876
 A:Experimental source: cultivar Columbia; BAC clone T22B4
 C:Genetics:
 A:Map position: 4
 A:Introns: 104/1; 186/3; 216/3; 275/1; 331/3; 373/3; 403/2
 A:Note: T22B4.100

T08191 Length: 415 February 11, 2000 15:52 Type: P Check: 2415 ..

1 MAPRAVRAP IGYYIVSYSS RSGSNEYST VASKETLSQ YKSSVPSGCT
 51 SLVRGFGNEI RSESEAPPA VSDQSLIKQ LRERTSAPIK DVKASLVCN
 101 WDLVGFEGFD TEAAOKDLK RGVLAASKS SRTAEGMLA VQNGKAVAV
 151 IELNCETDEV ARNEIFQYLV SPALSHWLL LNDGLFLAM AKHALVESS
 201 SQQVSGVFPE GPDLFEFEKL NLDHPRVNGE TTVSNAVTEV AAIMGVNKF
 251 RRGFLMSKSS AGVLAAYLT SPQAGIVSL EVEGNTQLE AIORVSELA
 301 MHVAAKPLF LSKDLVSSA MANREILKS QASTGNOM AIKIVYGL
 351 RKFEEVALM EQKFIVNDAI NIKTLVDNIS KEVGSPPAVT DELRVEGEC
 401 IERLEADEP VAQTA

11AA_SEQUENCE 1.0
 P1:T09662 - multicatalytic endopeptidase complex chain Y13 homolog - alfalfa
 (alfalfa)
 N:Alternate names: proteasome chain Y13 homolog
 C:Species: Medicago sativa (alfalfa)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
 C:Accession: T09662
 R:Frings, G.
 submitted to the EMBL Data Library, April 1996
 A:Description: Isolation of an alfalfa DnaU-like gene and a Y13 proteasome
 subunit homologous gene showing differential expression in embryogenic and
 non-embryogenic callus
 A:Reference number: Z16763
 A:Accession: T09662
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-101 <FRU>
 A:Cross-references: EMBL:Z71998; NID:e992787; PID:e239056
 A:Experimental source: variety Kangelander
 C:Keywords: hydrolase; proteinase

T09662 Length: 101 February 11, 2000 15:52 Type: P Check: 3889 ..

1 CDPSGNYGWM KAGAIANNQ AAOSILKODY KDDITREAV NLALVLSKI
 51 MDRTSLTSDK LELAELVLP SGRKXYOVS PENITKLIVK SGVTPARET
 101 R

11AA_SEQUENCE 1.0
 P1:S23774 - trisec phosphate/3-phosphoglycerate/phosphate translocator
 P:cursor - garden pea
 C:Species: Pisum sativum (garden pea)
 C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Sep-1997
 C:Accession: S23774; S23816; A37791; S23775; S13761
 R:Willege, D.; Fischer, K.; Wachter, E.; Link, T.A.; Fluegge, U.I.
 Planta 183: 451-461, 1991
 A:Title: Molecular cloning and structural analysis of the phosphate
 translocator from pea chloroplasts and its comparison to the spinach phosphate
 translocator.
 A:Reference number: S23774
 A:Accession: S23774
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA

A/Residues: 1-402 <WIL>
 A/Accession: S23816
 A/Molecule type: protein
 A/Residues: 73-95:97-102;104;106;108-115 <WIL1>
 A/Note: 227-leu was also found
 R.Schneil, D.J. Blodet, G.; Pain, D.
 J. Cell Biol. 111, 1825-1838, 1990
 A>Title: The chloroplast import receptor is an integral membrane protein of
 chloroplast envelope contact sites.
 A/Reference number: A37791; M01D:91035597
 A/Accession: A37791
 A/Molecule type: mRNA
 A/Residues: 1-402 <SCH>
 A/Cross-references: EMBL:X54639; NID:920690; PID:920691
 A/Accession: S23775
 A/Molecule type: protein
 A/Residues: 'X', 74-98:249-260;269-289;324-350 <SCH1>
 A/Note: this protein was identified as chloroplast import receptor
 C/Keywords: chloroplast; transmembrane protein
 E/1-72/Domain: transit peptide (chloroplast) #status experimental <TNP>
 E/73-402/Product: triose phosphate/3-phosphoglycerate/phosphate translocator
 #status experimental <MAT>
 E/97-117/Domain: transmembrane #status predicted <TM1>
 E/129-151/Domain: transmembrane #status predicted <TM2>
 E/168-190/Domain: transmembrane #status predicted <TM3>
 E/214-234/Domain: transmembrane #status predicted <TM4>
 E/273-292/Domain: transmembrane #status predicted <TM5>
 E/342-364/Domain: transmembrane #status predicted <TM6>
 E/367-388/Domain: transmembrane #status predicted <TM7>
 S23774 Length: 402 February 11, 2000 15:52 Type: P Check: 5889 ..

1 MESRVLSRAT TLSSLPITLKH LHRPLPLANS LPSVRSFGSV SDGGLTWGR
 51 OLAPELCSFY LKKGASLLRP CPATAGGNDG AGEKVAEFG FFSRYPALTT
 101 GFFFTWYFL NVIPNLLNK IYTFPYPF VSVTHLVGV VYGLVSWTVG
 151 LPRRAIDGN LKLLIPYAV CHALGHVTSN VSFAVAVSE THTVKALEPF
 201 FNAAAOPIFL GOSIPITLMT SLAPVYIGVS MASLTLSFN WGFSTAMIS
 251 NISFTYRSY SKRAMTDMDS TNYAYISII ALIYCIPPAL IIEGTLTKT
 301 GRNDALAKVG LVKFSDFLV VGMFYHLNQ VAITLTERVA PLTHAVGVNL
 351 KRVFVIGFSI IIFGNKISTQ TGIGTGIAIA GVALYSIKK QIEEKKRAK
 401 AA

!!AA_SEQUENCE 1.0
 P1:S22500 - embryonic protein ecp31 - carrot
 C/Species: Daucus carota (carrot)
 C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
 A/Accession: S22500
 R.Kiyosue, T.; Yamaguchi-Shinozaki, K.; Shinozaki, K.; Higashi, K.; Satoh, S.;
 Kamada, H.; Harada, H.
 Plant Mol. Biol. 19, 239-249, 1992
 A>Title: Isolation and characterization of a cDNA that encodes ECP31, an
 embryogenic-cell protein from carrot
 A/Reference number: S22500; M01D:92322951
 A/Accession: S22500
 A/Molecule type: mRNA
 A/Residues: 1-256 <KIY>
 A/Cross-references: EMBL:X60593; NID:918336; PID:918337
 C/Genetics:
 A/Genes: ecp31
 S22500 Length: 256 February 11, 2000 15:52 Type: P Check: 9347 ..

1 MSQQQRRRQ QEQPIKIGDV FDSVGLSSQ VPAPNDASAM QAAENNVLAK
 51 TQKGGPASYA QSAANLQR GVVPHEGTP IASEGVATIS EAEIATRIII

101 TEAVGGQVVG QYLEPGRFKK ESPAGVLGSD SITIGALEET TALTAGDKPV
 151 DOSDAAAIQA AEVRASGAY PGVAAAAS AADYNRTMNN VASRTKIGDV
 201 IADRSIRLAE DKAVTREDAE GVGAEEYRN PEMMTYPGV ASSMAAARL
 251 NODPTF

!!AA_SEQUENCE 1.0
 P1:T16984 - transcription factor homolog BRF3 - curled-leaved tobacco
 C/Species: Nicotiana glauca (curled-leaved tobacco)
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 A/Accession: T16984
 R.Borisjuk, N.V.
 submitted to the EMBL Data Library, October 1996
 A/Reference number: Z18621
 A/Accession: T16984
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-165 <BOR>
 A/Cross-references: EMBL:Y09106; NID:61008749; PID:e280517
 A/Experimental source: somatic embryo
 C/Keywords: transcription factor
 T16984 Length: 165 February 11, 2000 15:52 Type: P Check: 4925 ..

1 MNVEKLRKKA GSVRTGKGT MRRKKAVHK TTTTDRKLO STLKRIQVNA
 51 IPAIEVNIF KEDVVIQFIN PKVQASIAN TWVYSGSPQT KTLQGYSSN
 101 YSPVGDNIE SLREASRAVP ESRASANGA PEGALPOED DDDEVELVA
 151 GQTFEAGNSE ERHTS

!!AA_SEQUENCE 1.0
 P1:S47086 - pir7a protein - rice
 C/Species: Oryza sativa (rice)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
 A/Accession: S47086
 R.Reilmann, C.; Mauch, F.; Dudler, R.
 submitted to the EMBL Data Library, June 1994
 A/Description: Host-response of rice to resistance-inducing agents:
 characterisation of a gene specifically regulated by infiltration of
 Pseudomonas syringae pv. syringae.
 A/Reference number: S47085
 A/Accession: S47086
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-263 <REI>
 A/Cross-references: EMBL:Z34271; NID:9498743; PID:9498744
 C/Genetics:
 A/Introns: 122/3; 168/3
 S47086 Length: 263 February 11, 2000 15:52 Type: P Check: 1305 ..

1 MEDQGRFVE VAGLGYGMC WYRVVAALRA AGHRAMLM DM AAGAPARA
 51 DEVSLEIS RPLDVAVAA APGERLVVG HSLGLSLAL AMERFPDKVA
 101 AAYFLACMP AAGKHMGITL EEFMRIRKP FFMDSKTIIVL NTNQEPRTAV
 151 LLOPKLIAR LYNRPSPEDL TLATMLVRPG TNYIDDDIMK DELLTEGNY
 201 GSVKRVFLVA MDDASDDEM ORWTIDLSPG VEVELAGAD HNAKSKFRE
 251 LQDLRLRIA KYD

!!AA_SEQUENCE 1.0
 P1:S14884 - hypothetical protein 7 - yeast (Hansenula polymorpha)
 C/Species: Hansenula polymorpha
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997
 A/Accession: S14884

R:Krutillina, A.I.; Seragina, S.A.; Tikhomirova, L.P.; Kryukov, V.M.
 submitted to the EMBL Data Library, April 1991
 A:Description: Nucleotide sequence of Hansenula polymorpha DNA region
 complementing DAK.
 A:Accession: S14854
 A:Reference number: S14854
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-146 <KRU>
 A:Cross-references: EMBL:X58862; NID:92762; PID:92769

S14884 Length: 146 February 11, 2000 15:52 Type: P Check: 7397

1 MGGTSGGIYS IFISALAKSL KDREIQGGY EYTPPELIAS LKDALDSIYR
 51 YTRARAGDRT LIDALAPVE OFPAKSGDLN RANKACHEGA ESTRLAKKF
 101 GRASYSEEE FKPEAEAGGL PDPGAIGLAA LVDFAPAEYS KIGSNL

!!NA_SEQUENCE 1.0
 P1:S43448 - FUN54 protein precursor - yeast (Saccharomyces cerevisiae)
 N:Alternate names: probable membrane protein YAL007C
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
 C:Accession: S43448; S40896
 C:Date: 13-Jan-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
 R:Clark, M.W.; Kemp, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.;
 Delaney, S.; Ouellette, B.F.F.; Barton, A.B.; Kaback, D.B.; Bussey, H.
 Yeast 10, 535-541, 1994
 A:Title: sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
 the 42 kbp SP07-CENT-CD15 region.
 A:Reference number: S43441; MUID:95028152
 A:Accession: S43448
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-213 <CIA>
 A:Cross-references: EMBL:L22015; NID:91339990; PID:9349746; MIPS:YAL007C
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library,
 November 1993
 C:Genetics:
 A:Gene: FUN54
 A:Cross-references: MIPS:YAL007C
 A:Map position: 1L
 C:Keywords: glycoprotein; transmembrane protein
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-215/Product: FUN54 protein #status predicted <AMT>
 F:183-203/Domain: transmembrane #status predicted <TM>
 F:170-175/Binding site: carbohydrate (asn) (covalent) #status predicted

S43448 Length: 215 February 11, 2000 15:52 Type: P Check: 5386

1 MIKSTIALPS FFIYLIILAV NSVAASSYA PVALISPARS KECLYIDMVT
 51 EDDSLAVGYQ VLTGNFEID FDIAPDGSV ITSEKOKKIS DILKSFQVG
 101 KTFCEFSNNY GIALKVEIF LEKEKTLTDE HEADVNDMDI INNAVEEID
 151 RNLNITFTL NYLAREMRN MSTVNSTESR LTMLSILILI ITAVISIAYV
 201 LLIQLEFTRG OKNYV

!!NA_SEQUENCE 1.0
 P1:S19425 - hypothetical protein YCR015C - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 14-Nov-1997
 C:Accession: S19425
 R:Batat, D.; Jacq, C.; Perez, J.; Shu, Y.
 submitted to the Protein Sequence Database, March 1992
 A:Reference number: S19424
 A:Accession: S19425
 A:Molecule type: DNA
 A:Residues: 1-317 <HAT>
 A:Cross-references: EMBL:X59720; NID:91907116; PID:e264479; PID:91907158;
 MIPS:YCR015C

C:Genetics:
 A:Map position: 3R
 S19425 Length: 317 February 11, 2000 15:52 Type: P Check: 1969

1 MKTIIDDFD EITRVDFIC TIAPLYLLN PLIKPEWHF TKTYMDGKH
 51 KYKGTSLP LSSGVPTII SOSNFNLF A DELKYNHNR VVELNSVNEI
 101 TKQITFKSIS LQMKTFAND QNHDDCLRD GERTCSSYV KNEBDFYVL
 151 SINNSKEFIH EYIGDRRLKN SHIFCNLDRK VSDKCSOSYN GEFCRLLTG
 201 SDRKYLGEI LDKIDSGCKR EGNCSYWI GDBETDLISI LHPSTNGVLL
 251 INPOENPSKF IKTEITLIGI PKDKISFEA DNGPAWLQFC EKEGKGAYL
 301 VKSWDSLNDL IMQYTKM

!!NA_SEQUENCE 1.0
 P1:S70131 - hypothetical protein YDR273W - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein D9954.4
 C:Species: Saccharomyces cerevisiae
 C:Date: 24-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 12-Dec-1997
 C:Accession: S70131
 R:Le, T.
 submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of S. cerevisiae cosmid 9954.
 A:Reference number: S70124
 A:Accession: S70131
 A:Molecule type: DNA
 A:Residues: 1-365 <LET>
 A:Cross-references: EMBL:U51030; NID:91332633; PID:91230644; MIPS:YDR273W
 C:Genetics:
 A:Map position: 4R
 A:Note: YDR273W

S70131 Length: 365 February 11, 2000 15:52 Type: P Check: 1796

1 MGKRNKRGKE NNAKTSFLK VENIKNTNSG IELPSQDYTN VERKESSPKT
 51 DFPLITKEHV NRTKDSNILD YPTIGDIVSS VEKLVLMEL KIAPEYVDI
 101 LIRKILASQ GYLEPAFNSL LYSSPEENT DELAPMKRIS VEDYSKINVS
 151 EILQREIFDD IDEPFGGCI NGSMVISKIE SETSLIAEHI GNTSTPGSNR
 201 EVAESTRNVA VABGHNTIIS NEDSILKGE KGRKEERKRG EKGVNSLKG
 251 AAVKYVAKSL KNNRIPYTK RNEPSNNLFD VLNCDSEDE EEOQVETNTS
 301 NGERNKOGN TEVPEKORDS ADRLPAKDDG GYKSAFGTDS GCLFAADAKD
 351 EKKQVHPSRQ ELSEFT

!!NA_SEQUENCE 1.0
 P1:S69699 - hypothetical protein YDR415C - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 21-Nov-1997
 C:Accession: S69699
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, August 1995
 A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461,
 9831, and 9410.
 A:Reference number: S69553
 A:Accession: S69699
 A:Molecule type: DNA
 A:Residues: 1-374 <DIE>
 A:Cross-references: EMBL:U33007; NID:927685; PID:9277713; MIPS:YDR415C
 C:Genetics:
 A:Map position: 4R

S69699 Length: 374 February 11, 2000 15:52 Type: P Check: 3355

```
1 MROSLFVLF NVAIANSYP YEPRLVQNG ENEMVEPES EKLNRRRGY
51 KEFDVTKHTS FLPEFNKEE PTVPTYNRP EISKREYVD SIKNDKSGM
101 HKNLAKTSF YTRYKSDHG FESAEMLAAT IANTKDIDP DLTIEHPDH
151 KEMKOYSIV RVGSTTPEP IITIGSHDS INLLPSIMA APGADDNNG
201 TVTNMEALRL YTENFLKRGF RPNNTVEHF YSAEBSGLLG SLDFFTAYAK
251 QKHVRAMLQ QDMTGVSDP EDEHGIYTD YTPPALTDI KLINSLIST
301 PYRDTCCGYA CSDHGSATRN GFPGSEVIES EFKTKNKYH STMDTLDRLS
351 LAHMAERTKI VLGVIIEIGS WSAW

!!AA_SEQUENCE 1.0
P1:S50587 - hypothetical protein YER084w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 21-Nov-1997
R:Dietch, F.S.
Submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 9747, 8198, 9781, and
lambda clones 3612 and 6052.
A:Reference number: S50428
A:Accession: S50587
A:Molecule type: DNA
A:Residues: 1-128 <DIE>
A:Cross-references: EMBL:U18839; NID:9603313; PID:9603322; MIPS:YER084w
C:Genetics:
A:Map position: 5R

S50587 Length: 128 February 11, 2000 15:52 Type: P Check: 5798
1 MEELICTPY HSNLEMFLL FFCPSKRAAR GHPEFLETLG YKSNHLIKL
51 LPSLFTKRV MLNDSSHPPS PDFPTGSSAS PRVKLRPSTL WAPPLTVSSD
101 FRASSSTAP VVTDKFVTP AVSKRTQD

!!AA_SEQUENCE 1.0
P1:S64117 - hypothetical protein YGL109w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G3065
C:Species: Saccharomyces cerevisiae
C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 12-Dec-1997
R:Castagnoli, L.; Pauluzzi, S.; Minenkova, O.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64112
A:Accession: S64117
A:Molecule type: DNA
A:Residues: 1-107 <CAS>
A:Cross-references: EMBL:Z72631; NID:91322653; PID:e243985; PID:91322654;
MIPS:YGL109w
A:Experimental source: strain S288C
R:Castagnoli, L.; Pauluzzi, S.; Minenkova, O.
Submitted to the EMBL Data Library, April 1996
A:Reference number: S69421
A:Accession: S69421
A:Molecule type: DNA
A:Residues: 1-107 <CAW>
A:Cross-references: EMBL:X97644; NID:91310710; PID:e240104; PID:91310715
C:Genetics:
A:Map position: 7L

S64117 Length: 107 February 11, 2000 15:52 Type: P Check: 6897
1 MAQNPLADI QVRRYAKR RMEGCKKNC TIATISLQY YCRRSLSHKS
51 CFPFSQNAF SRPLPSESY ETWALELAF CLTRPYCTFH SLEISSQQL

101 TLRLPLG

!!AA_SEQUENCE 1.0
P1:S46801 - hypothetical protein YKL027w homolog - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YHR003c
C:Species: Saccharomyces cerevisiae
C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 12-Dec-1997
R:Favell, T.
Submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9780.
A:Reference number: S46794
A:Accession: S46801
A:Molecule type: DNA
A:Residues: 1-429 <FAV>
A:Cross-references: EMBL:U10555; NID:9500813; PID:9500821; MIPS:YHR003c
C:Genetics:
A:Map position: 8R

S46801 Length: 429 February 11, 2000 15:52 Type: P Check: 5569
1 MANNTWKLIA TTALISYFT QLAQSWKREY KLSQAANKRK TVSRPRQYDD
51 HLFREQLARN YAFLEGGRM KIKEQYIVY GAGEVGSWC TMLIRSGCQK
101 IMIDPENIS IDSLNTHCCA VLSDIGKPKV OCLKEHLSKI APMSERAKRA
151 KAMIKENSHD LIPADESPT FIVDCLDNLE SKVDLEVAH HNKIDVISM
201 GVATKSDPTR VSINDISMTF FDPISRCVRR KLRKRGATG ISVFSNEM
251 DRRDDILSP IDCEHRAINA VRDEALRHP ELGTMPGIFG LSIATWLTK
301 VSGYPMKENE VKNRLKFTYDS ILETFOKOMA RLENKRESS LGLGLEVGYI
351 VEEMFRGKSP ISGYSTKLAL TKWEANKEIS LNNVLMTRK EOEIHEKRIL
401 LDGEKLTAVY SEEVLDFTER LFKEEEYS

!!AA_SEQUENCE 1.0
P1:S64829 - hypothetical protein YLR007w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L1543
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 14-Nov-1997
R:Vandenbol, M.; Portetelle, D.; Hliger, F.
Submitted to the Protein Sequence Database, May 1996
A:Reference number: S64742
A:Accession: S64829
A:Molecule type: DNA
A:Residues: 1-336 <VAN>
A:Cross-references: EMBL:Z73179; NID:91360297; PID:e245489; PID:91360298;
MIPS:YLR007w
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 12R

S64829 Length: 336 February 11, 2000 15:52 Type: P Check: 7470
1 MEYHEBOVSA PTGDATAKX LLOYLSARG ICHENALLILA LMRLETDAST
51 LNTENSIOGW VDKINDYINA INVKLNLIGY KIRINHGIG RNAVTLAKQO
101 NESEFEDNTA IRAHNDYAV LOSIVLPESN RFPVYNLAS TEEKLATRF
151 NONLEIFMKW AIDOFMISGE TIVGPALET SIIVEVARI LVAATGDSNL
201 AKWRKSTFT VGSTNLFQPO ELTATDIED LRLCEIKWF YRTOEGAFGI
251 DLRCAIELEE YLTSWNLNT CONCHKLAQ GYRCNNEGR ENNEDETEGNS
301 LSOIHWDCF KHYTHVSKN CDRGSSLIT EGYVIY
```

11AA_SEQUENCE 1.0
P1:S68481 - hypothetical protein YLR165c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein 19632.11
C:Species: Saccharomyces cerevisiae
C>Date: 20-Jul-1996 #sequence_revision 23-Aug-1996 #text_change 12-Dec-1997
C:Accession: S68481
R:Vardoli M.
Submitted to the EMBL Data Library, July 1996
A:Description: The sequence of S. cerevisiae cosmid 9362.
A:Reference number: S68471
A:Accession: S68481
A:Molecule type: DNA
A:Residues: 1-254 <YNU>
A:Cross-references: EMBL:U51921; NID:g1234842; PID:g1234853; MIPS:YLR165c
C:Genetics:
A:Map position: 12R

S68481 Length: 254 February 11, 2000 15:52 Type: P Check: 4630 ..

1 MSLKQIPII PENTHYFIVN KPGIPSQP DCRTGRTHP NIDPTPLER
51 FPAITYSHRE VELCRIVYHRL DHCYVGMLI AKTRDSYVF SFLQKGGNN
101 GKRLQRTYA IVSSGKFNK PNNEYIKYGP KYNFLISHGG REITKFEVD
151 ENCIYQIVLT GKRLQINNV SOLNQPILN DKRGSTVNF PALFNDQIAL
201 HSACITIKIG LQTKHLIPM EHNNIGQLMS RKYVNEGEF TPIKEVILLE
251 NMDO

11AA_SEQUENCE 1.0
P1:S69848 - hypothetical protein YML101c-a - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 21-Nov-1997
C:Accession: S69848
R:Barrell, B.G.
Submitted to the EMBL Data Library, August 1994
A:Reference number: S47445
A:Accession: S69848
A:Molecule type: DNA
A:Residues: 1-105 <BAR>
A:Cross-references: EMBL:X80835; MIPS:YML101c-a
C:Genetics:
A:Map position: 13L

S69848 Length: 105 February 11, 2000 15:52 Type: P Check: 7663 ..

1 MYRIKIKNH KSTGVCMLR FLTLILQGI FTGSMFTIP PASRWFLAT
51 ALSSVSSGA CIAGSILDP NCVSYIDTY EQNPVDEIS RVEPSSDHAK
101 SYMGE

11AA_SEQUENCE 1.0
P1:S77699 - inner cell wall mannoprotein ICWP - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YLR390w-a
C:Species: Saccharomyces cerevisiae
C>Date: 13-Jan-1998 #sequence_revision 09-May-1997 #text_change 07-May-1999
C:Accession: S77699; S78080
R:Noukadir, I.; Armero, J.; Abad, A.; Santandreu, R.; Zueco, J.
J. Bacteriol. 179, 2154-2162, 1997
A:Title: Identification of a mannoprotein present in the inner layer of the cell wall of Saccharomyces cerevisiae
A:Reference number: S77699; MUID:97234625
A:Accession: S77699
A:Molecule type: DNA
A:Residues: 1-238 <MOU>
R:Du, Z.
Submitted to the EMBL Data Library, January 1995
A:Description: The sequence of S. cerevisiae cosmid 8084.
A:Reference number: S55944

A:Accession: S78080
A:Molecule type: DNA
A:Residues: 1-238 <DU>
A:Cross-references: EMBL:U19729; NID:g525097; PID:g2580460; MIPS:YLR390w-a
C:Genetics:
A:Gene: ICWP
A:Cross-references: MIPS:YLR390w-a
A:Map position: 12R
C:Keywords: cell wall

S77699 Length: 238 February 11, 2000 15:52 Type: P Check: 7370 ..

1 MRATLLSSV VSLALISKEV LATPPACILA CVAQVKSSS TCDSLNOVTC
51 YCEHENSARV KCLDISCPNN DADAISAKR SCSEQNASL GDSSSSASSS
101 ASSSSKASSS TRASSSSASS STRASSSSAS SSTKASSSSA APSSKASSST
151 ESSSSSSST KAPSSSESSS TYVSSSKOAS STSEAHSSSA ASSTVQGEIV
201 SSALPTSTAV ISTFSESGSN VLEAKSVFT AAVAAALI

11AA_SEQUENCE 1.0
P1:A40550 - peroxisomal membrane protein PAS3 - yeast (Saccharomyces cerevisiae)
N:Alternate names: PAS3 protein; protein YDR329c
C:Species: Saccharomyces cerevisiae
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Dec-1997
C:Accession: A40550; S59794
R:Kochfeld, J.; Veenhuis, M.; Kunau, W.H.
J. Cell Biol. 114, 1167-1178, 1991
A:Title: PAS3, a Saccharomyces cerevisiae gene encoding a peroxisomal integral membrane protein essential for peroxisome biogenesis.
A:Reference number: A40550; MUID:91373453
A:Accession: A40550
A:Molecule type: DNA
A:Residues: 1-441 <HOE>
A:Cross-references: GB:X58407; NID:g41102; PID:g41103
R:Du, Z.
Submitted to the EMBL Data Library, July 1995
A:Description: The sequence of S. cerevisiae cosmid 9798.
A:Reference number: S59783
A:Accession: S59794
A:Molecule type: DNA
A:Residues: 1-441 <DU>
A:Cross-references: EMBL:U32517; NID:g914989; PID:g915004; MIPS:YDR329c
A:Experimental source: strain S288C (AB972)
C:Genetics:
A:Gene: SGD:PEX3; PAS3
A:Cross-references: MIPS:YDR329c; SGD:S0002737
A:Map position: 4R
C:Keywords: peroxisome; transmembrane protein
C:Keywords: peroxisome; transmembrane #status predicted <TM1>
F:136-152/Domain: transmembrane #status predicted <TM2>

A40550 Length: 441 February 11, 2000 15:52 Type: P Check: 166 ..

1 MAPNRSRL LQHRGKVL I SLTGIALFT TGSVVVFYK RWLYKQOLRI
51 TEQHRIKQI KRFEQOTQD SLTYTYELP VWRNVINEND LNDLSIYOL
101 KQKQQLRA KSSERESP LKSKAEIWE LELKSLIKLV TVTYVSLI
151 LTRIQILNL TRNEYLDSAI KLIMQENCN KIONRFYVNV TSWMSDEPK
201 ADDAVYMAK KSKKEGEVY INEQAFISLS WMLINKGWS YNEITINOE
251 IEDGHPHD TLLEFPSSR LTNIFRNTS QIQOONNNL TSLILPKDS
301 GQEFILSQL DADALTSFHS NTLVNOVLN ELTQCIESIA TSIVLESILN
351 ESHFIMNV GIKTIARKP GQEDQOYOV AVFAMSKDC COEMLQTTAG
401 SSGSGVNEY LATDSVQPL DDLASVSVSN FGVSFSFSFK P

```

!!AA_SEQUENCE 1.0
P1:S67699 - probable membrane protein YDL151c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D1554
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence, revision 12-Jul-1996 #text_change 14-Nov-1997
R:Perez, J.; Blugson, C.; Delaveau, T.; Jacq, C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67693
A:Accession: S67693
A:Molecule type: DNA
A:Residues: 1-193 <PER>
A:Cross-references: EMBL:Z74199; NID:g1431236; PID:e253062; PID:g1431237;
MIPS:YDL151C
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 4L
C:Keywords: transmembrane protein
F:114-130/Domain: transmembrane #status predicted <TM>
S67699 Length: 193 February 11, 2000 15:52 Type: P Check: 4697

1 MSTSLFSLSS PSSSSSMRLR ASNSFPLNF FLDDAPSL SSSSASFSL
51 SAPLSIVSRP FCTRDDPLPS DLNPLKSP FSLKRPFRAN GPLEITCVLF
101 KYLAIRLCWP PAPVTLLEFL LKCFSLPFL DSSSFTLDA AASLSLDFL
151 AATFGINENE GLADPPLEE ESFNGKRF PLLILLINEC YPA

!!AA_SEQUENCE 1.0
P1:S60916 - probable membrane protein YNL266w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N0800
C:Species: Saccharomyces cerevisiae
C>Date: 15-Feb-1996 #sequence, revision 01-Mar-1996 #text_change 17-Mar-1999
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.M.
submitted to the EMBL data library, October 1995
A:Description: The sequence of a 24152 bp segment from the left arm of
chromosome XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
genes.
A:Reference number: S60909
A:Accession: S60916
A:Molecule type: DNA
A:Residues: 1-139 <SEN>
A:Cross-references: EMBL:X92494; NID:g1045236; PID:g1045244
R:Sen-Gupta, M.; Lyck, R.; Niedenthal, R.K.; Fleig, U.N.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63235
A:Accession: S63235
A:Molecule type: DNA
A:Residues: 1-139 <SEN>
A:Cross-references: EMBL:Z71542; NID:g1302322; PID:e239650; PID:g1302323;
MIPS:YNL266W
A:Experimental source: strain S288C
R:Sen-Gupta, M.; Lyck, R.; Fleig, U.; Niedenthal, R.K.; Hegemann, J.H.
A:Title: The sequence of a 24 152 bp segment from the left arm of chromosome
XIV from Saccharomyces cerevisiae between the BNI1 and the POL2 genes.
A:Reference number: S65111; MUID:96310631
A:Accession: S65118
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-139 <SEF>
A:Cross-references: EMBL:X92494; NID:g1045236; PID:g1045244
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October
1995
C:Genetics:
A:Map position: 14L
A:Note: YNL266w
C:Keywords: transmembrane protein
F:34-50/Domain: transmembrane #status predicted <TM>

F:64-80/Domain: transmembrane #status predicted <TM2>
S60916 Length: 139 February 11, 2000 15:52 Type: P Check: 7226

1 MMLNHTYKL LSYFLRKASN RFNSSSSSF SCSFLVLEFV VFPSDFFSI
51 TSFLISFGIL SFLIFSLFC LGFLTYIGL ASALSLSLS KAKIGSSSL
101 SSISPEGLK SEEMLEDED KERSSLLYET SYVAISFK

!!AA_SEQUENCE 1.0
P1:S66918 - probable membrane protein YOR044w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O2771
C:Species: Saccharomyces cerevisiae
C>Date: 12-Jul-1996 #sequence, revision 12-Jul-1996 #text_change 14-Nov-1997
R:Landt, O.; Hiesel, R.; Unseld, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66907
A:Accession: S66918
A:Molecule type: DNA
A:Residues: 1-157 <LAN>
A:Cross-references: EMBL:Z74952; NID:g1420170; PID:e251967; PID:g1420171;
MIPS:YOR044W
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 15R
C:Keywords: transmembrane protein
F:4-20/Domain: transmembrane #status predicted <TM1>
F:24-40/Domain: transmembrane #status predicted <TM2>
S66918 Length: 157 February 11, 2000 15:52 Type: P Check: 8726

1 MIBLELYVL LVIGSLGYIC RTCIAPFLIP FLGIYAFDLF LVYRNHILV
51 SQENFYRKRL GRSKTNRPB SPLRKIYSS GDCEDTLIGQ VDRLEFLIS
101 TISHSKRFF STRFQTKSGI NSAIKDANVE TTSVSSFTN LHLTSSERG
151 YTIAGSI

!!AA_SEQUENCE 1.0
P1:JT0589 - pectinesterase (EC 3.1.1.11) precursor - Aspergillus niger
N:Alternate names: pectin methyl-esterase
C:Species: Aspergillus niger
C>Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text_change 13-Sep-1998
R:Khanh, N.Q.; Rutkowski, E.; Leidinger, K.; Albrecht, H.; Gottschalk, M.
Gene 106, 71-77, 1991
A:Title: Characterization and expression of a genomic pectin methyl
esterase-encoding gene in Aspergillus niger.
A:Reference number: JT0589; MUID:92039066
A:Accession: JT0589
A:Molecule type: DNA
A:Residues: 1-331 <KHL>
A:Cross-references: EMBL:X54145; NID:92372; PID:92373
A:Experimental source: strain RH5344
A:Note: the authors translated the codon GAG for residue 148 as Ile
R:Khanh, N.Q.; Albrecht, H.; Rutkowski, E.; Loeffler, F.; Gottschalk, M.;
Jany, K.D
Nucleic Acids Res. 18, 4262, 1990
A:Title: Nucleotide and derived amino acid sequence of a pectinesterase cDNA
isolated from Aspergillus niger strain RH 5344.
A:Reference number: S10487; MUID:90332436
A:Accession: S10487
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-331 <KN2>
A:Cross-references: EMBL:X52902; NID:92368; PID:92369
R:Markovic, O.; Joernvall, H.
Protein Seq. Data Anal. 3, 513-515, 1990
A:Title: Tomato and Aspergillus niger pectinesterases. Correlation of
differences in existing reports: large species variations.

```

A:Reference number: S13959; MUID:91212339
 A:Accession: S13959
 A:Molecule type: protein
 A:Residues: 18,'XG','21-32','F','34 <MAR>
 A:Experimental source: strain 8 WX-41
 C:Comment: This enzyme belongs to the group of pectic enzymes and catalyzes the hydrolysis of methyl ester groups of galacturonic acid residues of pectin.
 C:Genetics:
 A:Gene: pmea
 A:Introns: 39/3; 85/2; 112/1; 161/1; 181/2; 246/1
 A:Keywords: carboxylic ester hydrolase; glycoprotein; hydrolase
 F:1-17/Domain: signal sequence #status predicted <Sig>
 F:18-33/Product: pectinesterase #status experimental <MAT>
 F:95,283/Product: binding site: carbohydrate (asn) (covalent) #status predicted
 J0589 Length: 331 February 11, 2000 15:52 Type: P Check: 609 ..

1 MKSILASVL FAATALAASR MTPASGAIV AKSGGDYDTI SAAVDALSTT
 51 STETOTIFIE EGSYDEQVYI PALSGKLIVY GQEDITTYT SMLVITHAI
 101 ALADVNDDE TATLRNTAEG SAITYNLINIAN TCGGACHQAL AVSAYASEOG
 151 YVACQFTGYQ DTLAETGYQ VYAGTYIEGA VDFIFGQAR AMFHECDIRY
 201 LEPSSASIT ANGRSSESD SYVIRKSTV AADGNDVSS GTYILGRPMS
 251 QYARVCFORT SMTDVINHLG WTEMSTSTPN TENVTEVYG NTGTGAEGR
 301 ANFSSEITEP ITISWLGSD WEDWVDISYI N

11AA_SEQUENCE 1.0
 P1:S03833 - hypothetical protein 1 - chestnut blight fungus
 C:Species: Cryphonectria parasitica, Endothia parasitica (chestnut blight fungus)
 C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Sep-1997
 C:Accession: S03833
 R:Rae, B.P.; Hillman, B.L.; Tartaglia, J.; Nuss, D.L.
 EMBL J. 8, 657-663, 1989
 A:Title: Characterization of double-stranded RNA genetic elements associated with biological control of chestnut blight: organization of terminal domains and identification of gene products
 A:Reference number: S03833; MUID:82931594
 A:Accession: S03833
 A:Molecule type: DNA
 A:Residues: 1-319 <RAE>
 A:Cross-references: EMBL:X14524; NID:92624; PID:92625
 A>Note: the authors translated the codon CAG for residue 156 as Gly
 S03833 Length: 319 February 11, 2000 15:52 Type: P Check: 3269 ..

1 MAOLRRPSQS LVISESYDPT TVDPFVSVRT EEVVPACIT LMERYDSGCD
 51 VGPPLSHDL RRLRTPDGVK KCOVHELEPT VLKSGSTGV PEHPVLAIF
 101 IGRPRCSLE ORTEKELDSRF LQLVHGGLPA RPSYMIARP RPYRGICSSR
 151 NSLSAQFGG YCYLSAIVDS ARMRVARTG WCVNAVADYLR LIQWGRNSF
 201 GSQJIEKSAV DRYVHVAVDA EYQSEDDGAL FYQAILGLAE KQPLARIGR
 251 LNLIAEFAP GSALRVEPVT PQVTRRKGST EMTGRDPTV FRWQNGYGH
 301 QHPCSCGYG VEROFRPS

11AA_SEQUENCE 1.0
 P1:A34051 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 29/21K chain precursor
 - Neurospora crassa
 N:Alternate names: complex I dehydrogenase 29/21K chain
 C:Species: Neurospora crassa
 C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Sep-1997
 C:Accession: A34051
 R:Videla, A.; Iroeschug, M.; Werner, S.

Biochem. Biophys. Res. Commun. 166, 280-285, 1990
 A:Title: Primary structure, in vitro expression and import into mitochondria of a 29/21-kDa subunit of complex I from Neurospora crassa.
 A:Reference number: A34051; MUID:90147712
 A:Accession: A34051
 A:Molecule type: mRNA
 A:Residues: 1-201 <YND>
 A:Cross-references: GB:M32244; NID:9168778; PID:9168779
 C:Keywords: mitochondrion; NAD; oxidoreductase
 A34051 Length: 201 February 11, 2000 15:52 Type: P Check: 7750 ..

1 MAAKYVTVG KTTAGGVVPY SOKTYVOSYG VMERIRARA IDPKRNSGVP
 51 LVPYRNKSP GSUPLDLYDD PVTIAGDIA DNYWKRDRR RNPPLSYVG
 101 QAEVALLSV GSATHPRVEL VGENGSKQV AADACKTGS LAKYEGTGV
 151 EAGKVLAEI GTLEPLRSGE KLEGGKWDV YKQLAEPFS YEAVPCRSF
 201 S

11AA_SEQUENCE 1.0
 P1:S16556 - hypothetical protein 1 - Podospora anserina
 C:Species: Podospora anserina
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
 C:Accession: S16556
 R:Turcq, B.; Delau, C.; Denayrolles, M.; Begueret, J.
 Mol. Gen. Genet. 228, 265-269, 1991
 A:Title: Two allelic genes responsible for vegetative incompatibility in the fungus Podospora anserina are not essential for cell viability.
 A:Reference number: S16556; MUID:91360075
 A:Accession: S16556
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-289 <TUR>
 S16556 Length: 289 February 11, 2000 15:52 Type: P Check: 4238 ..

1 MSEPFGIVAG ALNVAGLFNN CVDFEYVQL GRPFGDYER COLRLDIARA
 51 RLKRWGAIVK INDDPRFHS APTKSYQLA KSIVEILLIS FESAOKTSKR
 101 YELVADQDL VPEDKDKMP IGRALHRLN DLVSRROKOT SLAKTAMAL
 151 YDGSLEKIV DOVARYVEL EKAPIEAVC HKLAIEIDE VEDASLTIL
 201 KDAAGCIDAA MSDAAAKID AIVGRNSAKD IREERARVQ LGNVYTAAL
 251 HGRIRISDOT TNSVETVYVK GESRVLYGNE YGKKGWMDN

11AA_SEQUENCE 1.0
 P1:S16557 - hypothetical protein 2 - Podospora anserina
 C:Species: Podospora anserina
 C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
 C:Accession: S16557
 R:Turcq, B.; Delau, C.; Denayrolles, M.; Begueret, J.
 Mol. Gen. Genet. 228, 265-269, 1991
 A:Title: Two allelic genes responsible for vegetative incompatibility in the fungus Podospora anserina are not essential for cell viability.
 A:Reference number: S16557; MUID:91360075
 A:Accession: S16557
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-289 <TUR>
 S16557 Length: 289 February 11, 2000 15:52 Type: P Check: 4187 ..

1 MSEPFEIVAG ALVAGLFNN CVACFEYVQL GRHFGDYER COLRLDIARY
 51 RLKRWGEAVQ INDDPRFHS APIKSYOLA KSIVEILLIS FESAOKTSKR
 101 YELVADQDL VFEEDKDKMP IGRALHRLK DLVSRROKOT SLAKTAMAL

151 YGKSLKIV DOVAGFVDEL EKAFPIEAVC HKLAENEIEE VEDEASLITL
201 KDAAGGIDAA MSDAAAKRID AIYGRNSAKD IRTEKARVQ LGNVYTAARL
251 HGEIRISDOT TNSVETVYVKR GSKVLIENE YGKRGFMDN

!!AA_SEQUENCE 1.0
P1:JN0451 - phosphoribosylanthranilate isomerase (EC 5.3.1.24) - fungus
(Filobasidium floriforme)
N:Alternate names: PRA1
C:Species: Filobasidiella neoformans, Cryptococcus neoformans
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 10-Oct-1997
R:Perfect, J.R.; Rude, T.H.; Penning, L.M.; Johnson, S.A.
A:Title: Cloning the Cryptococcus neoformans TRP1 gene by complementation in
Saccharomyces cerevisiae
A:Reference number: JN0451; MID:93083987
A:Accession: JN0451
A:Molecule type: DNA
A:Residues: 1-312 <PER>
A:Cross-references: GB:M74901; NID:9167395; PID:9167396
C:Genetics:
A:Gene: TRP1
C:Superfamily: trpF homology
C:Keywords: Intramolecular oxidoreductase; isomerase
F:77-312/Domain: trpF homology <TRF>

JN0451 Length: 312 February 11, 2000 15:52 Type: P Check: 2342

1 MSTSIVNNA LNRDGVVICA LSGISSHEDV EKYKEGVKG VIVGELMKA
51 SOTKATLRL IGLPPELVYS KPRPLVYKIG IRSTNDAKLA INAGADILGY
101 ILVPGTKRCI STTAREISA LVQASARSOS SKPLEPSLSS PMFTQSALL
151 SSRKRPDLVG VFONOSLSDI LSAVDEIGLD LVQIHGDEPO AVAKTIPVAV
201 VAVFVSESEG IVRGELIRP GLNQAILLDA GAGSGGGES KAFPEHAKR
251 LIQSGEVSE GHVPLPVILA GGLTPENVQO ALNRIVKARG VMMSVSGSE
301 REGSRRLRL RS

!!AA_SEQUENCE 1.0
P1:B44418 - surface antigen - Trypanosoma brucei
C:Species: Trypanosoma brucei
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997
R:Accession: B44418
R:Jackson, D.G.; Smith, D.K.; Luo, C.; Elliott, J.F.
J: Biol. Chem. 268, 1894-1900, 1993
A:Title: Cloning of a novel surface antigen from the insect stages of
Trypanosoma brucei by expression in COS cells.
A:Reference number: A44418; MID:93131939
A:Contents: 427 NITAT 1.1
A:Accession: B44418
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-412 <TAC>
A:Note: sequence inconsistent with the nucleotide translation
C:Keywords: surface antigen

B44418 Length: 412 February 11, 2000 15:52 Type: P Check: 3830

1 MCIEOLVHSD GEFKTAIVAC LCCLLIGSPV LIGGVLFSL SDDPRDNFK
51 AVSAPRPKL ESWTGTSDV KATVRROSLS VAGIGSIPSY YTEATPVASG
101 NNDGSLVYK VNITVAPFT RRSPLHATRE RMFSCSSSOC SGYSRRCDOQ
151 EKHEDFRNKC YSEGGYRSTQ SSKRLGKRC GYCKOEYVLS KLYLVAASDG

201 KGEYESTQY OSALYSFGL SQYEAVPOD KYOVOLYSEG DPTALERET
251 MGEFGEVFN RTMGIAICIA GSLILLEIA VCVCVVCCL KRGSSSNDT
301 GDPDPQGDG SPYTYGQSGP PPPRGYAYGQ PLPQGGTYG QPPPOGAT
351 YGQPPPPQG YPYGQPPPPQ QGGTYTKPP PQGGTYGQP PPPQGGHXY
401 GQACCPQNP TV

!!AA_SEQUENCE 1.0
P1:T14622 - hypothetical protein - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
R:Accession: T14622
R:Anderson, B.; Aslund, L.; Pettersson, U.
A:Submitted to the EMBL Data Library, March 1998
A:Description: 93.4 Kd of complete sequence from chromosome 3 of Trypanosoma
cruzi
A:Reference number: 218159
A:Accession: T14622
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-425 <AND>
A:Cross-references: EMBL:AF052833; NID:93063554; PID:93063555; PIDN:AA014090.1
C:Genetics:
A:Map Position: 3

T14622 Length: 425 February 11, 2000 15:52 Type: P Check: 9305

1 MSRLNPATLE SIMOVGLVG RSPFPSSSG KARESGGRIS TSDRLRLKA
51 VIIEPALSTI VGGMIIPFLV VEKTTGLR RMIAMRDKN RDDYEHAAP
101 LSHISHYLP VMAEAASCDL LKTSFFQVSL PRETRHLFC RVEGTLVEL
151 TRLPGYKAS PEILQITSA IAGTVVHR LCAPPPIVI VWIAGNIRIT
201 GSKSGATLME AGLRNPDGC HASMGEEES GATOTTFLLV RPDHTRAVS
251 LSDKFSVSR AMPALNSSTI AEMEVSAR LYAAAILGTR SCQYEFIRA
301 VRRRLALNR GIVQETSPN VPPRVLGE RLRHIENIT VSESSPFRK
351 HRPPSSMRH SMDGEPLFQ TPATIKLPEE NGRSLFLSC RPRRAVLAL
401 SAFRHLAIH HSLGQHFA ARSGE

!!AA_SEQUENCE 1.0
P1:T15458 - hypothetical protein C08A9.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
R:Accession: T15458
R:latreille, P.
A:Submitted to the EMBL Data Library, December 1995
A:Description: The sequence of C. elegans cosmid C08A9.
A:Reference number: 218353
A:Accession: T15458
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-331 <LAT>
A:Cross-references: EMBL:U42844; NID:91125797; PID:91125804; PIDN:AA053819.1;
GSPDB:G00028; CESP:C08A9.7
C:Experimental source: strain Bristol N2; clone C08A9
A:Gene: CESP:C08A9.7
A:Map Position: 10
A:Introns: 100/1; 188/2; 247/3

T15458 Length: 331 February 11, 2000 15:52 Type: P Check: 2648

1 MVSATRVRR SSTTSATQ ORTPSLMPA SPFITMDEVL EKENRETVN

```

51 ASKDIANKKL ALTLLEIYPE MMRGGMVA KKKOAFGAM YRRTGKIYRC
P1:T15482 - hypothetical protein C10A4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15482
R:Pauley, A.
Submitted to the EMBL Data Library, March 1995
A:Description: The sequence of C. elegans cosmid C10A4.
A:Reference number: Z18358
A:Accession: T15482
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <RND>
A:Cross-references: EMBL:U23454; NID:g733580; PID:g733583; PIDN:AAC46517.1;
CESP:C10A4.3
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C10A4.3
A:Introns: 31/1; 59/3; 85/2

T15482 Length: 217 February 11, 2000 15:52 Type: P Check: 121 ..

1 MARMIRSFRI ILMCSFISI CLYTAMTE SSINRIIPR TQCLTCNPHQ
51 EEPYQILNHL SOISEFTKIN FAYDLCLFVI TTFNLVKDQ SGADDELISI
101 GVFCIKYILIP YMLYNFSGA HSPAHNISTYL LNNEVCGTIG NCLMGNNVARI
151 MYSCISYVII MMLAFPALIA SMGWYIITIV RTIDQPLPMV TEQDDLPAY
201 SVLFPSQQL LKKEDEC

11AA_SEQUENCE 1.0
P1:T15182 - hypothetical protein C18E3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15182
R:Connell, M.; Magg, L.
Submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C18E3.
A:Reference number: Z18304
A:Accession: T15182
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-150 <CON>
A:Cross-references: EMBL:AF000265; NID:g1947147; PID:g1947151; PIDN:AAB52944.1;
GSPDB:GN00019; CESP:C18E3.4
A:Experimental source: strain Bristol N2; clone C18E3
C:Genetics:
A:Gene: CESP:C18E3.4
A:Map position: 1
A:Introns: 33/3; 104/2; 123/3

T15182 Length: 150 February 11, 2000 15:52 Type: P Check: 9118 ..

1 MSLTMDKCS RMKTIYVVT TTTTLDLHY YGAEVLSY VSVVSIQNTQ
51 QFVVYDIKAL LYINCNDLIP KASKLYITIIQ TLPLPSRNL IROPKILTF
101 ISOINVDKOC NSPMFLKYF SPSKLLKFIK NFVQSMNTF QKRIIYIHH

11AA_SEQUENCE 1.0
P1:T15715 - hypothetical protein C30G12.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15715
R:Latreille, P.
Submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid C30G12.
A:Reference number: Z18393
A:Accession: T15715
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-280 <LAT>
A:Cross-references: EMBL:U21319; NID:g687832; PID:g687834; PIDN:AAC46671.1;
CESP:C30G12.2
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C30G12.2
A:Introns: 184/3; 199/3

T15715 Length: 280 February 11, 2000 15:52 Type: P Check: 4709 ..

1 MADVADPEVE SRPQRPFKNI MITGANRGIG FGLVKHFLY DCIELLITC
51 RNPEKADELN ALKNDRLHY IALNVDDDES IKKYFDEVSS LVSSGLNML
101 INNAGILLY EVDGPKICKR TMKOLETNS VSAVILTOIF LPLITASA
151 AEQDEASIDR ASINISSTM ASIENNNGCF DGPKRFFHY FLVNSQEMTT
201 AYMSKSALN AFARQFREL SKYHILVTSF CPGWRTDMG GDNADLVNE
251 STKLSANIL RLDNNNGLY FDRFLHIPN

11AA_SEQUENCE 1.0
P1:T15729 - hypothetical protein C31H1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15729
R:Le, T.
Submitted to the EMBL Data Library, December 1995
A:Description: The sequence of C. elegans cosmid C31H1.
A:Reference number: Z18395
A:Accession: T15729
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <LET>
A:Cross-references: EMBL:U42848; NID:g1125817; PID:g1125822; PIDN:AA63610.1;
CESP:C31H1.6
C:Genetics:
A:Gene: CESP:C31H1.6
A:Introns: 5/1; 38/3; 77/3; 106/3; 215/3; 234/1; 271/3

T15729 Length: 304 February 11, 2000 15:52 Type: P Check: 770 ..

1 MSDECSISQ KSPAFSPEE KMLKSNRANE POSFVKGDA AKSDILDLDE
51 VYTNARKTAE DVTKSKAEGK IKKRYLSYNI NMTKLFQEGS EPTVITIGAK
101 PNWVDAKOE KVIRKTFITN ITORQEXSF KTERCTRSSS TVIIEGYCR
151 GAVSIAKLT PCEVEANAG FTEVNLNHI GENTSEBELT WGVDSQVAVP
201 PGELYALVY IIEDIVTRDF RIENRLSGV LVITINLKEN NSLVITIEGN
251 ICDIITGMPD YTAGFRFDG PAIVETRGO CIFRFEGIEOK VRINEFSINS
301 SRRY

11AA_SEQUENCE 1.0
P1:T16486 - hypothetical protein F56D3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans

```

Mon Feb 14 08:07:32 2000

pir.cat

Page 172

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T16486
R:Wilcox, L.

Submitted to the EMBL Data Library, June 1995

C/Species: Caenorhabditis elegans cosmid F56D3.

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T16486

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-374 <WILL>

A/Cross-references: EMBL:U029942; NID:9861330; PID:9861332; PIDN:AA68355.1;

CESP:F56D3.1

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:F56D3.1

A/Introns: 21/2; 41/1; 121/2; 150/2; 245/2; 315/3; 338/3

T16486 Length: 374 February 11, 2000 15:52 Type: P Check: 1289

1 MKRLILIFM LGVKAQEEV KTVERRASE ADQNGEKFKI YGTIDGVR

51 ELNEVDGGO IEVTPDNTPN VEDPFGTYD NNVVSADNEK ILFVPRKPYRP

101 GGDKEPKRT TOAPTTIRTT RPPATIRI IPFVOTTPAT FATLPAPAV

151 TRADIQFPV QIVATIPSF VRPOVPTAP RVIAPOQPOQ PFRIOPIOP

201 QOPRPPRFQ PRPOAFTOA PFTOAPFTQA PFRPPFRQ PFRAPRQAP

251 QPAPTPAPR FPPOGONLR SGQCOASIFY ISTPISGPR LITFRAVAN

301 TVDQCATCH EFNCASCNPB SGHRKSGKF HESIDKSTE VKQPRHREK

351 IDEVTSERS SOLNLKOLE DLDD

!!AA_SEQUENCE 1.0
P1:T16528 - hypothetical protein K02F3.4 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T16528

R:Bentley, D.

Submitted to the EMBL Data Library, May 1994

A/Description: The sequence of C. elegans cosmid K02F3.

A/Reference number: 218530

A/Accession: T16528

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-419 <BER>

A/Cross-references: EMBL:U00052; NID:9485125; PID:9485133; PIDN:AA50709.1;

CESP:K02F3.4

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:K02F3.4

A/Introns: 72/1; 101/1; 114/3; 220/1; 327/1

T16528 Length: 419 February 11, 2000 15:52 Type: P Check: 3866

1 MGQVCESSL IFWMESETH IPLHTSFTLL VGCVGRRLTL VLFHHTLSL

51 ILRLHTVSCR PSSVSDWYT SSRQFOGPT TATVRAQIYT SSSLSFSP

101 VNSVASWTRP SMEPLSAPLA VGFSAMPSP PLPLASHL QDAVHADHNH

151 RLINSCRL PDLTTVPQ LHOINSAQS PLQFTHQPC DLIVYDECT

201 TYTTMTRRS LPRADRYLSE PNMSTPQASH GSFEVDVEY TSFDYAFDP

251 PVGEAFPPAA NSSASAPDA DFLDLSLMS CDITSLDIYI HDLSIASPSN

301 ASHISPPASD IDPVDEFPQ LVNNKIQSS SSVYETTTT RGRKTSVSS

351 DSSSDYRHRK DKNNLASQKS ROKROAKIRE SKREERLEK RYQOLQAWL

401 TLETQVEDYK RLYMAYVR

!!AA_SEQUENCE 1.0
P1:S40941 - hypothetical protein ZK632.9 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 06-Jan-1995

C/Accession: S40941

R:Beiers, M.

Submitted to the EMBL Data Library, February 1993

A/Reference number: S40933

A/Accession: S40941

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-219 <BER>

A/Cross-references: EMBL:222181

C/Genetics:

A/Introns: 11/3; 56/2; 94/1; 152/1

S40941 Length: 219 February 11, 2000 15:52 Type: P Check: 8579

1 MAAFEQOTGG PDLTGSGR TKSDRYEHKH ASQEGGOTRK VQOTASNGEA

51 KRKEKWKMS RPLGFIGYEF GDDMEVQOM IEKSNMEOA KMLEQCKML

101 ECTETMPEES EPVPMKCLDF EEAFOSESYS KGYESPYKNI SPLEKDAVTV

151 NTMSHCPADO IAKLIRNQN SVYTLGIEBA RQCRKGLN VLKPTGSASP

201 RYLQTPPKN VAEETGSO

!!AA_SEQUENCE 1.0
P1:S44857 - PAR2.1 protein - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 31-Oct-1997

C/Accession: S44857

R:Wilson, R.

Submitted to the EMBL Data Library, March 1994

A/Description: The sequence of C. elegans plasmid PAR2.

A/Reference number: S44857

A/Accession: S44857

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-170 <WILL>

A/Cross-references: EMBL:U00025; NID:9458884; PID:9458885

C/Genetics:

A/Introns: 31/3; 53/2

S44857 Length: 170 February 11, 2000 15:52 Type: P Check: 5825

1 MRLSISTSK STVRGMSILTS KMAEOPSKQ EVDLFAKRP OHNHPORRH

51 AYSVKNVELV GGVALDPLRK TGRNGRPYLI FNIIINSYK QODGTLLDOT

101 ERHNVSVFGR QAEILSTIKR KGSRLWQGR LHSYGOKDE QGNRTORNTY

151 IIAQTVQPLA RAARENPDQH

!!AA_SEQUENCE 1.0
P1:S44865 - R05D3.9 protein - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997

C/Accession: S44865

R:Wilson, R.

Submitted to the EMBL Data Library, December 1992

A/Description: Sequence of the C. elegans cosmid R05D3.

A/Reference number: S44732

A/Accession: S44865

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-336 <WILL>

A/Cross-references: EMBL:L07144; NID:9156412; PID:9156417

C/Genetics:

A/Introns: 46/2; 113/2; 153/3; 262/1; 302/3

Mon Feb 14 08:07:32 2000

pl.r.cat

Page 173

```
S44865 Length: 336 February 11, 2000 15:52 Type: P Check: 399
1 MTLLEVRAS TKLTITILSR IEMGRILEKP RILGTWSEFR IGLNCESLSG
51 AIQRLPNESK LSEFKKGIWG GKRRSVATIG VFPSDDOTSF QHFOIAEKL
101 SKRYVAVFI NENSKPALIT YHAHKQRT DYSGKDDPAT LMEFTKSI
151 PSIIISNGF TTDLLEHQKR PLIIIGAG EFEPFSLSA RODAKTYIF
201 TKIDSESEMI KTKKALGVG ENESITVPLN KDRVHRIPIS KRKCDHLOK
251 ILOMITSEA DOVLSTKEPH PLRYLOKEV NEVGFEEETL VLPDHTLFD
301 SDPFSRHPI TEGGGGCPF MOGGAPTES SQHSEL

!!AA.SQUENCE 1.0
P1:S06613 - chorion protein sl9 - fruit fly (Drosophila grimshawi)
C:Species: Drosophila grimshawi
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998
C:Accession: S06613
R:Kertner-Cruzado, J.C.; Swimmer, C.; Fenerjian, M.G.; Kafatos, F.C.
Genetics 119: 663-677, 1998
A:Title: Evolution of the autosomal chorion locus in Drosophila. I. General
organization of the locus and sequence comparisons of genes sl5 and sl9 in
evolutionarily distant species.
A:Reference number: S06612; MUID:88297142
A:Accession: S06613
A:Molecule type: DNA
A:Residues: 1196 <MAP>
A:Cross-references: GB:X53422; EMBL:X12635; NID:97419; PID:97422
C:Genetics:
A:Gene: sl9-1
A:Cross-references: FlyBase:FBgn0012320
A:Map position: 5
A:introns: 5/3

S06613 Length: 196 February 11, 2000 15:52 Type: P Check: 4664
1 NMTFATLAI ISACLAVSGC GYGSPICYG GPINAGLRNV SSIGQSGDG
51 AAASMAASG GDNQPEVETIA GGAPRYSSO NLRPILLNSG YHGGLNDWIG
101 RIAQIVGGGR SLGGHLGGHL GGHLCGRIGG NYGGRITRRR FLYQPAAGTL
151 LYPGONSRYR ISSPEYSKY ILPRAAPV AKLIIPONTY GSQVY

!!AA.SQUENCE 1.0
P1:S69241 - Dreg-5 protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 16-Apr-1999
C:Accession: S69241
R:Van Gelder, R.N.; Krasnow, M.A.
EMBO J. 15: 1625-1631, 1996
A:Title: A novel circadianly expressed Drosophila melanogaster gene dependent
on the period gene for its rhythmic expression.
A:Reference number: S69241; MUID:96203080
A:Accession: S69241
A:Molecule type: mRNA
A:Residues: 1298 <YAN>
A:Cross-references: GB:U65105; GB:S81693; NID:91513135; PID:91513136
C:Genetics:
A:Gene: Dreg-5

S69241 Length: 298 February 11, 2000 15:52 Type: P Check: 7903
1 MTLAKYILA CCLGAFHIO ISSSSAIPW EFLTRNEKMS HLXSTFAQLV
51 SVCKSTRAY GGLPVNQCK NLIGYSAKL OTLSVQOLA LDYQORDANE
101 LWSIIGGDH PSGASIVTTR QPLQDPLTP PASLITLIR QOLPHGASHA

151 HPLOSSGSAT NPFESEGEQ HKYAMDNDKA YQGPQSSSE LPVAAALISE
201 PSKRFLTCPL VIRVRPDGSP VEEDKMPLP RDEDLPYLS WSGRRAQQA
251 PODRHNDLK AVALRLHPE RPPATPPDAE ASVPATGVR SRSEDDQA

!!AA.SQUENCE 1.0
P1:A4254 - ecdysone-induced protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change 16-Feb-1997
C:Accession: A4254
R:Cherbas, L.; Schultz, R.A.; Koehler, M.M.D.; Savakis, C.; Cherbas, P.
J. Mol. Biol. 189: 617-631, 1986
A:Title: Structure of the Eip28/29 gene, an ecdysone-inducible gene from
Drosophila.
A:Reference number: A4254; MUID:87060956
A:Accession: A4254
A:Molecule type: DNA
A:Residues: 1-255 <CHB>
C:Genetics:
A:Gene: FlyBase:Eip71CD
A:Cross-references: FlyBase:FBgn0000565
A:introns: 17/3; 82/3; 169/3
C:Keywords: alternative splicing

A4254 Length: 255 February 11, 2000 15:52 Type: P Check: 68
1 MSLLITSVY HPEIKDLSTV RNEOKELNIS PYHDVNTKA TATFGMGFW
51 GASIVGATR GVLRITVGYA GGSSDLPTFR KMSDTEVLE IDYDPAIVSF
101 KELLDLFVNN HEYGLTPIK ROYASLLIYH DEOKQVANA SKLEODERA
151 PEITTEIAS KENFYPAEAY HOKYRLQGRK DLASLNLSP KLIQISYVAT
201 KLNQYLAGV GIEQKAEAE TTGSDAHPPA VLLPRAER GPGSLILTP
251 NVNHR

!!AA.SQUENCE 1.0
P1:S78042 - Ig mu chain C region, membrane-bond form - channel catfish
(Fragment)
C:Species: Ictalurus punctatus (channel catfish)
A:Variety: Channel catfish
C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 24-Oct-1998
C:Accession: S78042
R:Wilson, M.R.; Marcuz, A.; van Ginkel, F.; Miller, N.W.; Clem, L.W.;
Middleton, D.; Warr, G.W.
Nucleic Acids Res. 19: 5227-5233, 1990
A:Title: The immunoglobulin M heavy chain constant region gene of the channel
catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces the
membrane form of the molecule.
A:Reference number: S12833; MUID:90384824
A:Accession: S78042
A:Molecule type: DNA
A:Residues: 1-351 <WIL>
A:Cross-references: EMBL:X52617
A:Experimental source: erythrocyte
C:Genetics:
A:introns: 101/3; 203/3; 304/3; 349/3
C:Keywords: alternative splicing; glycoprotein; heterotrimer; immunoglobulin;
transmembrane protein

S78042 Length: 351 February 11, 2000 15:52 Type: P Check: 1000
1 VQSAFSLFP VMOCGSADG LVTLGCVTRD LASADGLSFI WPDAGSALT
51 DVQYPAVQA TGGYTSVSHV RVKASDMGN KFTCEYKNG LSGKASLQK
101 PYERELHASL ILTPTQTEI DNGTATFVCL ATPSPKSHT EKWTLEKIDI
151 SNKVENIVS QNKGNFTALS VLELSASEWT SSTSPKCEF QQXNNHVEKE
```

Mon Feb 14 08:07:32 2000

pl.r.cat

Page 174

201 ASYAPDPTKO POKYITGPST EDILIRRAQ LECRAEGDTG FRSIKWLIGN
251 REISSISNLS SKTVSLQTH IGFEWINGT EFICEVEHEA FTQOYEKVF
301 KRENVCINWST EHFHEMEND DDNMANTALI FVFLFLLTF YSIGVTVFV
351 K

11AA_SEQUENCE 1.0
P1:S69288 - early lactation protein precursor - brush-tailed possum
C/Species: Trichosurus vulpecula (brush-tailed possum)
C/Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C/Accession: J05761
R/Plotte, C.P.; Grigor, M.R.
Arch. Biochem. Biophys. 330, 59-64, 1996
A/Title: A novel marsupial protein expressed by the mammary gland only during
the early lactation and related to the Kunitz proteinase inhibitors.
A/Reference number: S69288; MUID:96230240
A/Accession: S69288
A/Molecule type: mRNA
A/Residues: 1-102 <PRO>
A/Cross-references: EMBL:U34208; NID:q1565292; PID:q1002806
A/Accession: S74274
A/Molecule type: Protein
A/Residues: 21-30 <PRT>
C/Superfamily: animal Kunitz-type proteinase inhibitor homology
F:1-20/Domain: signal sequence #status Predicted <SIG>
F:21-102/Product: early lactation protein #status Experimental <MAT>
F:43-93/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
S69288 Length: 102 February 11, 2000 15:52 Type: P Check: 5524

1 MKTITATCL ALSLVGMSS EKLIDIRAN SLENLSRLVP SLCLPSGRG
51 NCOSQILRF YVATSHCEV FLYSGGNG NNFDSLECL KTCRLNRYN
101 NN
11AA_SEQUENCE 1.0
P1:T14143 - ATPase subunit 8 - orangutan mitochondrion (SGC1)
C/Species: mitochondrion Pongo pygmaeus (orangutan)
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T14143
R/Boral, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
A/Title: Recent African origin of modern humans revealed by complete sequences
of hominoid mitochondrial DNAs.
A/Reference number: 159384; MUID:95132634
A/Accession: T14143
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-68 <HOR>
A/Cross-references: EMBL:D38115; NID:d1027246; PID:d1007887; PIDN:BA07309.1
C/Genetics:
A/Genome: mitochondrion
A/Genetic code: SGC1
C/Keywords: mitochondrion
T14143 Length: 68 February 11, 2000 15:52 Type: P Check: 3550

1 MPOINTTTL TVITPILAL FLITQKLIN SHLHPTPK FTKTRPHAK
51 WELKTKIYS PHSLPPOS

11AA_SEQUENCE 1.0
P1:JC5761 - cytokine-inducible SH2 protein 3 - human
C/Species: Homo sapiens (man)
C/Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 07-May-1999
C/Accession: JC5761
R/Masuhara, M.; Sakamoto, H.; Matsumoto, A.; Suzuki, R.; Yasukawa, H.; Mitsui, K.; Wakikawa, T.; Iannamura, S.; Sasaki, A.; Misawa, H.; Yokouchi, M.; Ohtsuda, M.; Yoshimura, A.
Biochem. Biophys. Res. Commun. 239, 439-446, 1997

A/Title: Cloning and characterization of novel cis family genes.
A/Reference number: JC5760; MUID:98008857
A/Accession: JC5761
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-225 <MAS>
A/Cross-references: DDBJ:AB006967; NID:q2463522; PID:d1023405; PID:q2463523
C/Comment: This protein plays a role in the negative regulation of cytokine
signaling by interacting with specific targets.
C/Superfamily: SH2 homology
F:45-144/Domain: SH2 homology <SH2>
JC5761 Length: 225 February 11, 2000 15:52 Type: P Check: 447

1 MYHSKRPAA GMSRPLDLS RLKTFSSKE YOLVNAVVR LQESGEWSA
51 VTGEANULL SAEPAITFLI RDSSDORHFF TISVKTOSGT KNLRIQCEG
101 SFSIQSDPRS TQVPRFDCV LKLVHMYMP QGAPSFSP TEPSSVPEQ
151 PSAPPLPQSP PRRAIYISG GEXIPLVLSR PLSSNVATLQ HLCRTVNGH
201 LDSYKATQL PGPREFLDQ YDAPL

11AA_SEQUENCE 1.0
P1:S43928 - estradiol 17beta-dehydrogenase (EC 1.1.1.62) type 3 - human
N/Alternate names: 17-beta-hydroxysteroid dehydrogenase
C/Species: Homo sapiens (man)
C/Date: 10-May-1995 #sequence_revision 10-Nov-1995 #text_change 26-May-1999
C/Accession: S43928
R/Beissler, W.M.; Davis, D.L.; Wu, L.; Bradshaw, K.D.; Patel, S.; Mendonca, B.B.; Elliston, K.O.; Wilson, J.D.; Russell, D.W.; Andersson, S.
Nature Genet. 7, 34-39, 1994
A/Title: Male pseudohermaphroditism caused by mutations of testicular
17-beta-hydroxysteroid dehydrogenase 3.
A/Reference number: S43928; MUID:94355972
A/Accession: S43928
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-310 <GEI>
A/Cross-references: GB:U05659; NID:q531161; PIDN:AA050066.1; PID:q531162
C/Genetics:
A/Gene: GDB:HSD17B3
A/Cross-references: GDB:347487; OMIM:264300
A/Map position: 9q22-9q22
C/Superfamily: short-chain alcohol dehydrogenase homology
F:49-229/Domain: short-chain alcohol dehydrogenase homology <SADH>
S43928 Length: 310 February 11, 2000 15:52 Type: P Check: 3434

1 MGVLDGEFFI LGLVAVLAC LAKCVFSSRC VLVNWKVP KSFLRSMGOW
51 AVITGADGDI GRAYSFEIAR RGLNVVLSR TLEKLEIAT EIBRTGRSV
101 KIIGADFTKD DIYEHKEKL AGLEIGILVN NVGMLPNLIP SHFLNAPDEI
151 QSLHCHITS YVKATQILIK HMESROGLI LNISSGIALF PMPLYMYSAA
201 SKAFVCAFSS ALQEFYRAKE VTIQVLPYRA VSIAMTYLN TNYITKTADE
251 FVKEISLNYVT IGGETCGCLA HEIILAGFLSL IPAWAFSGA FORLLITHYV
301 AYKLNATKVR

11AA_SEQUENCE 1.0
P1:J0175 - Frazzled protein-1b - human
C/Species: Homo sapiens (man)
C/Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C/Accession: J0175
R/Hu, E.; Zhu, Y.; Fredrickson, T.; Barnes, M.; Keisell, D.; Beeley, L.; Brooks, D.
Biochem. Biophys. Res. Commun. 247, 287-293, 1998

A:Title: Tissue restricted expression of two human frzbs in preadipocytes and pancreas.
 A:Reference number: JE0174; MUID:98308108
 A:Accession: JE0175
 A:Molecule type: mRNA
 A:Residues: 1-317 <HUA>
 A:Genetics:
 A:Gene: hFRP-1b
 A:Map position: 5q14.3-q12.1
 JE0175 Length: 317 February 11, 2000 15:52 Type: P Check: 3945 ..

1 MRAAAAGV RMAALLLG ALHAPARCE EHYHGMQAE PLHGRSTSKP
 51 PGCDIPADL PLCHTVGYKR MRLPILHEH SLAEVROAS SMLPLAKRC
 101 HSDTOVFICS LFAPVCLDRP IYPCRSLEA VRAGCAPLME AYGPWPBML
 151 HCHKEPLDND LCIAVCGHL PATAPPVTKI CAOCMEHSA DGLMEQMCSS
 201 DFVYKMKRIE IKIENGDRKL IGAOKKKLL KPGLKKEKI KRLVLMKNG
 251 AGCPCPQLDS LAGSELYMGR KYDGGQLLMA YRWMDKKNKE MKFAVKMFS
 301 YPCSLTYPEF YGAAPRH

11AA_SEQUENCE 1.0
 FI:177373 - gene N-ras protein - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 28-Feb-1997
 C:Accession: I77373
 R:Kiba, T.; Tsuda, H.; Pairololi, C.; Inoue, S.; Sugimura, T.; Hirohashi, S.
 Mol. Carcinog. 8, 312-318, 1993
 A:Title: Mutations of the p53 tumor suppressor gene and the ras gene family in intrahepatic cholangiocellular carcinomas in Japan and Thailand.
 A:Reference number: I57484; MUID:94107479
 A:Accession: I77373
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-22 <RES>
 A:Cross-references: GB:S68581; NID:9544858; PID:9544859
 A:Genetics:
 A:Gene: N-ras
 I77373 Length: 22 February 11, 2000 15:52 Type: P Check: 9167 ..

1 ETCULDTLL LTRRLPLCIC HQ

11AA_SEQUENCE 1.0
 FI:T14788 - hypothetical protein DKFZP564A122.1 - human
 C:Species: Homo sapiens (man)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14788
 R:Ucclelehoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 Submitted to the Protein Sequence Database, August 1999
 A:Reference number: Z18178
 A:Accession: T14788
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-112 <DUF>
 A:Cross-references: EMBL:AL110269
 A:Experimental source: fetal brain; clone DKFZP564A122
 A:Genetics:
 A:Note: DKFZP564A122.1
 T14788 Length: 112 February 11, 2000 15:52 Type: P Check: 6935 ..

1 MTGCGDYLLYLL AMMCMCLQAL LMKRWAFWR IDEDPYQVA PVSNGLCACAF
 51 STDGSVLAAG THDGSVYFWA TPPOVPSLOH LCRMSIRRYM PTOFVCLPFI
 101 PSKLEFLSY RI

11AA_SEQUENCE 1.0
 FI:T08722 - hypothetical protein DKFZP566D133.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
 C:Accession: T08722
 R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 Submitted to the Protein Sequence Database, May 1999
 A:Reference number: Z16468
 A:Accession: T08722
 A:Molecule type: mRNA
 A:Residues: 1-381 <KOE>
 A:Cross-references: EMBL:AL050050
 A:Experimental source: fetal kidney; clone DKFZP566D133
 A:Genetics:
 A:Note: DKFZP566D133.1
 T08722 Length: 381 February 11, 2000 15:52 Type: P Check: 6844 ..

1 RWTEQKFR EYPTWDTIR DEEDALDEL QYIGVTSPEC LORTGISLNI
 51 PAPQVPCISE KOESDVINAI LKQHTKEEF VEKHFNDLM KAVQODEPIF
 101 OKPOGAFYCC RLILSLGNM SMDKRSEHL LKINEXLKE LRNDSPQCR
 151 ETHKIAVEYV AEGQEDKHSI LTNNGSOAY EDFVAGLME VNLNHCQFM
 201 GGLQKKKSTG LTPYFATST VEVIFHMSTR MPSSDSDSLT KKLRLGNDL
 251 VHIWSEHTR DYRRGIIPTE PGDVLIVIP MKNMHSIOI MKKPEVPFEG
 301 PLFDGAIYNG KVLPIVYAT AINASRALKS LIPLYNFYE EKARLOTIV
 351 QHLEPTFE DFAQVSPA PYHHPSCAD H

11AA_SEQUENCE 1.0
 FI:T00702 - hypothetical protein F25965.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 01-Feb-1999
 C:Accession: T00702
 R:Lamerdin, J.E.; McCreedy, P.M.; Adamson, A.W.; Burkhardt-Schultz, K.; Garcia, R.; Kyle, A.; Ramirez, M.; Stillwagen, S.; Ganes, J.; Dangnan, L.; Bruce, R.; Quan, G.; Montgomery, M.; Oy, D.; Kobayashi, A.; Olsen, A.O.; Carrano, A.V.
 Submitted to the EMBL data library, October 1997
 A:Description: Sequence analysis of a 1mb region in 19q13.1.
 A:Reference number: Z14199
 A:Accession: T00702
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-196 <LAM>
 A:Cross-references: EMBL:AC002398; NID:92529398; PID:92477514
 A:Genetics:
 A:Map position: 19
 A:Note: intron positions not resolved (incomplete sequence)
 A:Note: F25965.1
 T00702 Length: 196 February 11, 2000 15:52 Type: P Check: 4343 ..

1 SWRWSKPGIN WMLSCSVWR RVWTGWSVM RKLGKHPQRP TINTYVILF
 51 DRSVDLAQFS ENTPLPYICR AMRNSPSYR ERECSRSSPL PPLPDEEGS
 101 EYTNKSRDY YKLPPTPPG PGDACSRI PSPLQPMQGS TPDDPSRPE
 151 PSPSTLIYRN MQRKRIQOR WRKASHRNQL RYSESKILIR EYTERO

11AA_SEQUENCE 1.0
 FI:A4478 - probable cell growth or differentiation regulator (alternatively spliced type I transcript) - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
 C:Accession: A4478
 R:Wetzel, J.N.; Kasperczyk, A.; Mohan, C.; Kronliris, T.G.
 Genomics 14, 309-319, 1992

A>Title: The HRA1 gene cluster: two upstream regions recognizing transcripts and a third encoding a gene with a leucine zipper domain.
 A:Reference number: A44478; M01D:93052330
 A:Accession: A44478
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-373 <ME>
 A:Cross-references: GB:M91083; NID:g184389; PID:g184390
 A>Note: sequence extracted from NCBI backbone (NCBIP:117869)

A44478 Length: 373 February 11, 2000 15:52 Type: P Check: 3379

```

1 MLGLAMEL KVVWDGIQRY VGVSEQITC QEVVIALQA IGGTGRFVLY
51 ORLEKEROL LPOECFVGAQ ATCGFASDV QFVLRTGPS LAGRPSSDSC
101 PPRPCLIRA SLVPRRAL GCEPRKTLTP EPAPLSRPG PAMPVPTPG
151 CCTDLRGLEL RVORNAEELR HEAFWEOELR REQAREREGQ ARLOALSAAT
201 AEHAARLOAL DAQARALEAE LQLAAPRGP PSPMASATER LHODLAVOER
251 QSAEVOGSIA IYSRALEAAE RALQAQOEL EELNRELKQC NLOQFTQOTG
301 ALPPPRPD RGPPTQGPL PPARESILG APSESHAGQ PPRGSPHDA
351 ELLEVAAPA PEMCLANQF QAL

```

!!AA_SEQUENCE 1.0
 P1:B44478 - probable cell growth or differentiation regulator (alternatively spliced type II transcript) - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Mar-1995
 C:Accession: B44478
 R:Wetzel, J.N.; Kasperczyk, A.; Mohan, C.; Krontiris, T.G.
 A>Title: The HRA1 gene cluster: two upstream regions recognizing transcripts and a third encoding a gene with a leucine zipper domain.
 A:Reference number: A44478; M01D:93052330
 A:Accession: B44478
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-373 <ME>
 A>Note: sequence extracted from NCBI backbone (NCBIP:117871)

B44478 Length: 337 February 11, 2000 15:52 Type: P Check: 908

```

1 MLGLAMEL KVVWDGIQRY VGVSEQITC QEVVIALQA IGGTGRFVLY
51 ORLEKEROL LPOECFVGAQ ATCGFASDV QFVLRTGPS LAGRPSSDSC
101 PPRPCLIRA SLVPRRAL GCEPRKTLTP EPAPLSRPG PAMPVPTPG
151 CCTDLRGLEL RVORNAEELR HEAFWEOELR REQAREREGQ ARLOALSAAT
201 AEHAARLOAL DAQARALEAE LQLAAPRGP PSPMASATER LHODLAVOER
251 QSAEVOGSIA IYSRALEAAE RALQAQOEL EELNRELKQC NLOQFTQOTG
301 ALPPPRPD RGPPTQGPL PPARESILG APSESHAGQ PPRGSPHDA

```

!!AA_SEQUENCE 1.0
 P1:A61188 - probable transcription factor SC1 - human
 C:Species: Homo sapiens (man)
 C:Date: 12-May-1994 #sequence_revision 09-Sep-1994 #text_change 07-May-1999
 C:Accession: A61188
 R:Ku, D.H.; Chang, C.; Konieczki, J.; Cammizaro, L.A.; Boghosian-Sell, L.; Alder, H.; Baserga, R.
 A>Title: A new growth-regulated complementary DNA with the sequence of a putative trans-activating factor.
 A:Reference number: A61188; M01D:91329275
 A:Accession: A61188

A:Molecule type: mRNA
 A:Residues: 1-358 <KUA>
 A>Note: authors translated the codon TAC for residue 19 as Thr
 C:Genetics:
 A:Gene: SC1
 A:Map position: 6p21-22

A61188 Length: 358 February 11, 2000 15:52 Type: P Check: 916

```

1 MLPCFOLLRI GGGRGGLTYT FHRPAGAGT YRLHRADLC DVALRPOEP
51 GLISGIAEL HAEPRGDWR VSLDHSSQG TLVNNVLRP GHRLEISGD
101 LITFGEPP GTPSEFYEM FQNVKYPQD PAITTPSRR GEARVAGFR
151 PMLPSGAPQ RPLSTSPAP KATLLNSIG SLKLRPQPL TTPSWGGPK
201 SLVPAPGE VGTTPSAPQ RNRKSVHRV LAELDESEP PENPPVLM
251 PRKLRVKA PLTPGNRG RPKTPVSAP MAPPAVAGT VOLVAAAPR
301 KRWGFSVM AVTSGKMPV LAASRLPGR PISDAQAGL AFSLRTRKA
351 PSDTRAH

```

!!AA_SEQUENCE 1.0
 F1:S63679 - signal transducer and activator of transcription 2 type a - human
 C:Species: Homo sapiens (man)
 C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 28-May-1999
 C:Accession: S63679
 R:Stojanovic, T.; Nishio, Y.; Kishimoto, T.; Akira, S.
 A>Title: Identification of alternative splicing form of Stat2.
 A:Reference number: S63679; M01D:96176320
 A:Accession: S63679
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-65 <SVG>
 A:Cross-references: GB:S81491; NID:g1478435; PID:NAA36225.1; PID:g1478436

S63679 Length: 65 February 11, 2000 15:52 Type: P Check: 1615

```

1 GFLRFSES SEGCTCSWV EHDDDKVLI YSVQPTKEY LOSLPTEIRI
51 IRHYQLTEE NIPEN

```

!!AA_SEQUENCE 1.0
 P1:A41795 - somatostatin receptor 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 24-Sep-1998
 C:Accession: A41795
 R:Yamada, Y.; Post, S.R.; Wang, X.; Tager, H.S.; Bell, G.I.; Saito, S.
 A>Title: Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney.
 A:Reference number: A41795; M01D:92108031
 A:Accession: A41795
 A:Molecule type: DNA
 A:Residues: 1-391 <YAM>
 A:Cross-references: GB:M81829; NID:g307433; PID:g307434
 A>Note: sequence extracted from NCBI Backbone (NCBIP:74767, NCBIP:74768)
 C:Genetics:
 A:Gene: GDS:STR1
 A:Cross-references: GDS:134185; OMIM:182451
 A:Map position: 14q13-14q13
 A:Introns: #Status absent
 A:Keywords: G protein-coupled receptor; glycoprotein; hormone receptor; lipoprotein; phosphoprotein; thiolester bond; transmembrane protein
 F1:84/Domain: transmembrane #status predicted <TM1>
 F1:95-120/Domain: transmembrane #status predicted <TM2>
 F1:132-153/Domain: transmembrane #status predicted <TM3>

F:173-195/Domain: transmembrane #status predicted <TM4>
F:220-250/Domain: transmembrane #status predicted <TM5>
F:269-296/Domain: transmembrane #status predicted <TM6>
F:302-336/Domain: transmembrane #status predicted <TM7>
F:4.44.48.381/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:130-208/Disulfide bonds: #status predicted
F:172/Binding site: phosphate (Thr) (covalent) (by CAMP-dependent kinase)
#status predicted
F:265/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase)
#status predicted
F:339/Binding site: palmitate (Cys) (covalent) #status predicted
A41795 Length: 391 February 11, 2000 15:52 Type: P Check: 8652 ..

1 MPPNQTASPP SSSPSPSPGS CGEGGSGRNP GAGADGME PERNASNGT
51 LSEGGGSAVL ISFTSYVCL VGLGNSMVI YTLIRAKMK TATNIYIILN
101 ATADELLMLS VPELYTSTLL RHPFGALIC RLYLSDAVN METSIYCLTV
151 LSVDRVAVV HPKARVRR PTVAKVNLG VWYLSLIVL PIVESRTAA
201 NSDGVACNM LMPEPAQRL VGFVLYTFIL GELLVGAIC LCVYLIAM
251 RMVALKAGW QKRSEKIT LVMWVVMVF VICWPFYVY QLVNFAEOD
301 DATVSOLSVI LGYANSCANP ILYGFLSDNF KRSFORILCL SWMDNAEER
351 VDIYATALKS RAYVEDFOP ENLESGVFR NGTCRSRTT L

!!AA_SEQUENCE 1.0
P1:JC5627 - STAT induced STAT inhibitor 3 - human
C:Species: Homo sapiens (man)
C>Date: 14-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 07-May-1999
C:Accession: JC5627
R:Matsumoto, S.; Ikegami, K.; Ueno, K.; Narazaki, M.; Naka, T.; Yamamoto, H.; Matsunoto, T.; Saito, H.; Hosoe, S.; Kishimoto, T.
Biochem. Biophys. Res. Commun. 237, 79-83, 1997
A:Title: Cloning and functional analysis of new members of STAT induced STAT inhibitor (SSI) family: SSI-2 and SSI-3.
A:Reference number: JC5626; MUID:97410118
A:Accession: JC5627
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MIN>
C:Comment: This protein plays a role in negative feedback control of Janus kinase-signal transduction and activator of transcription signaling pathway.
C:Superfamily: SH2 homology
F:45-144/Domain: SH2 homology <SH2>
F:189-197/Region: SC motif 1
F:210-218/Region: SC motif 2
JC5627 Length: 225 February 11, 2000 15:52 Type: P Check: 178 ..

1 MYTHSKFPAA GMSRPDTSI RLKTFSSKSE YQLVYVAVRK LOESGFYWSA
51 VTGGGANILL SARPAGTFEL RDSDQRHFE ALVKTQSGT KMLRIQCEGG
101 SFSLSDDPRS TQPYRPDCV LKLVHYMP PGARSPSP TPSSSEVEDO
151 PSAQPLPGSP PRRAIYISG GKIPVLVSR PLSSNVALIQ HLCRTVNGH
201 LDSYEKVTOL PGPIREFLDQ YDAPL

!!AA_SEQUENCE 1.0
P1:S41063 - translocon-associated protein beta chain - human
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 28-May-1999
C:Accession: S41063
R:Bodesscot, M.; Brisson, O.
Biochim. Biophys. Acta 1217, 101-102, 1994
A:Title: Cloning and sequence analysis of the beta subunit of the human translocon-associated protein.

A:Reference number: S41063; MUID:94114564
A:Accession: S41063
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-183 <BOD>
A:Cross-references: GB:X74104; NID:9452756; PIDN:CAA52207.1; PID:9452757
S41063 Length: 183 February 11, 2000 15:52 Type: P Check: 8093 ..

1 MRLSLFVLA LEAVTQAEFG ARLIAKSL NRYNVEGRLL TLQNIYNG
51 SSALDVELS DDFPPEDFG IVSGMLNWK DRIAPASNV HTVYLRPLKA
101 GFNFTSATI TYLAQEDGPV VIGSTAPQ GGLAQRERD RRFSPHLDW
151 AAFGVMTLPS IGIPILLWYS SKRYDTPKT KKN

!!AA_SEQUENCE 1.0
P1:A36679 - signal sequence receptor beta chain precursor - dog
N:Alternate names: glycoprotein 25H
C:Species: Canis lupus familiaris (dog)
C>Date: 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change 10-Sep-1997
C:Accession: A36679; B37273; S15510
R:Goerlich, D.; Prehn, S.; Hartmann, E.; Herz, J.; Otto, A.; Kraft, R.; Wiedmann, M.; Kneipel, S.; Döberstein, B.; Rapoport, T.A.
J. Cell Biol. 111, 2283-2294, 1990
A:Title: The signal sequence receptor has a second subunit and is part of a translocation complex in the endoplasmic reticulum as probed by bifunctional reagents
A:Reference number: A36679; MUID:9115924
A:Accession: A36679
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-183 <GOE>
A:Cross-references: GB:X53529; NID:9936; PID:9937
R:Wada, T.; Rindes, D.; Cameron, P.H.; Ou, W.J.; Doherty II, J.J.; Louvard, D.; Bell, A.W.; Dignard, D.; Thomas, D.V.; Bergeron, J.J.M.
J. Biol. Chem. 266, 19599-19610, 1991
A:Title: SRalpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane
A:Reference number: A37273; MUID:92011761
A:Accession: B37273
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-183 <WAD>
A:Cross-references: EMBL:X53591; NID:9845; PID:9846
C:Keywords: endoplasmic reticulum; glycoprotein; transmembrane protein
A36679 Length: 183 February 11, 2000 15:52 Type: P Check: 8392 ..

1 MRLASVLA LEAVSHAERG ARLIAKSL NRYNVEGRLL TLQNIYNG
51 SSALDVELS DDFPPEDFG IVSGMLNWK DRIAPASNV HTVYLRPLKA
101 GFNFTSATV TYLAQEDGPV VIGFTAPQ GGLAQRERD RRFSPHLDW
151 AAFGVMTLPS IGIPILLWYS SKRYDTPKS KKN

!!AA_SEQUENCE 1.0
P1:A56446 - Ig heavy chain V region (3H-3H scFv) - mouse (strain BALB/c)
C:Species: Mus musculus (house mouse)
C>Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 16-Aug-1996
C:Accession: A56446
R:Tang, F.M.; Politz, L.A.; Mahoney, W.C.; Schueler, P.A.
J. Biol. Chem. 270, 7829-7835, 1995
A:Title: A high affinity disoxin-binding protein displayed on M13 is functionally identical to the native protein.
A:Reference number: A56446; MUID:95225583
A:Accession: A56446
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-268 <TAN>
A:Cross-references: GB:U20617

Mon Feb 14 08:07:32 2000

pir.cat

Page 178

C:Keywords: heterotetramer; immunoglobulin

A56446 Length: 268 February 11, 2000 15:52 Type: P Check: 3609 ..

1 MAOVKLOESG AELVKPGASV KLSCTTSGFN IKDTYHWK ORPEQGLEMI
51 GRIPANCIT KYDPKFGKA TLADTSSNT AYLOLSLTS EDTAVYCAS
101 YLIRYENYW GGGTYTVSS GGGSGGSGS GGGSDIELT OSPALMSASL
151 GERYTMSGRA SSSVNFYIYW OOKSPASPKL WYVYTHLP GPAPAFSGG
201 GSNYSILITIS SMEGEDAATY YCOQFTSPF TFGSGTKLEI KRSAHHHH
251 HGAABOKLIS EEDLNGAA

!!AA_SEQUENCE 1.0
P1:B61188 - sc1 protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 19-Mar-1997 #sequence_revision 13-Mar-1998 #text_change 07-May-1999
C:Accession: B61188
R:Ku, D.H.; Chang, C.; Koniecki, J.; Canizazaro, L.A.; Boghosian-Sell, L.;
Alder, H.; Baseraga, R.
Cell Growth Differ. 2, 179-186, 1991
A:Title: A new growth-regulated complementary DNA with the sequence of a
putative trans-activating factor
A:Reference number: A61188; MUID:91329275
A:Accession: B61188
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-265 <KUA>

B61188 Length: 265 February 11, 2000 15:52 Type: P Check: 9720 ..

1 MUPCFLQIMT GGGKGVYIT FQAPAKSGCT YRLCCRADLC DYALRPOQEP
51 GLISGHAEL HAEIQGDDMR VSLIEDSSOG TLVNNVLP GRHLEISDD
101 LITFGQOGA GTSSSEFFY MFOQVVKPO DFAITVPRS KEGAGGFGP
151 MUPGPOPR LSTLSSAPKA TLILSMGSL SKLQAOPLF CGGGRPGKL
201 AIPSGHGA QVSPAPQPN RRSVHKYLA EMLDEVSPE SFLSVLEPR
251 KRLRVEKAL IASGE

!!AA_SEQUENCE 1.0
P1:S63681 - signal transducer and activator of transcription 2 type a - mouse
(fragment)
C:Species: Mus musculus (house mouse)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 29-Jan-1999
C:Accession: S63681
R:Sugiyama, T.; Nishio, Y.; Kishimoto, T.; Akira, S.
FEBS Lett. 381, 191-194, 1996
A:Title: Identification of alternative splicing form of Stat2.
A:Reference number: S63679; MUID:96176320
A:Accession: S63681
A:Status: preliminary; nucleic acid sequence not shown; not compared with
conceptual translation
A:Molecule type: mRNA
A:Residues: 1-64 <STUG>
A:Cross-references: GB:S61342; NID:91478438; PID:91478439

S63681 Length: 64 February 11, 2000 15:52 Type: P Check: 840 ..

1 GANFLRSEF SEGITCSMV EHODDKEI YVQPTREV LQSLPLEII
51 RHQVLAEN IPEN

!!AA_SEQUENCE 1.0
P1:C41795 - somatostatin receptor 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1993 #sequence_revision 31-Dec-1993 #text_change 10-Sep-1997

C:Accession: C41795
R:Hamada, Y.; Post, S.R.; Wang, K.; Tager, H.S.; Bell, G.I.; Seino, S.
Proc. Natl. Acad. Sci. U.S.A. 89, 251-255, 1992
A:Title: Cloning and functional characterization of a family of human and mouse
somatostatin receptors expressed in brain, gastrointestinal tract, and kidney.
A:Reference number: A41795; MUID:92108031
A:Accession: C41795
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-391 <YAM>
A:Cross-references: GB:M81831; NID:9201058; PID:9201059
C:Keywords: G protein-coupled receptor; hormone receptor; transmembrane protein

C41795 Length: 391 February 11, 2000 15:52 Type: P Check: 8110 ..

1 MFPNGTASP SSSPSPSGS CGEGACSRG GSGADGME PGRNASQGT
51 LSEOGSAIL ISFIYSVCL VGLGNSMYI YILIRYAKM TATNIYILNL
101 AINDELIMS VFELTSTLL RHPFGALLC RLVSVDVN MFTSIYCLTV
151 LSVDRVAVV HPIKARRR PLYAKVYNG VWVLSLVIL PIVFSRTAA
201 NSDGTACMK LMPEAOBWL VGFVYTFELM GELLVGAIC LCVYLIIAM
251 RMVALAGWQ QRKSEKIT LMYMYVYF VICMPPYV QLVNFAEOD
301 DATVSQLSVI LGVNSCANP ILYGFLDNF KRSFORIICL SWMDNAEEP
351 VDIYATLKS RAYVEDFOP ENLESGVFR NGTCASRIST L

!!AA_SEQUENCE 1.0
P1:S66270 - kynurenine-oxoglutarate transaminase (EC 2.6.1.7) /
asparagine-oxo-acid transaminase (EC 2.6.1.14) K precursor - rat
N:Alternate names: kynurenine aminotransferase / glutamine transaminase K
N:Contains: asparagine-oxo-acid transaminase (EC 2.6.1.14) K, cytosolic;
asparagine-oxo-acid transaminase (EC 2.6.1.14) K, mitochondrial; cysteine
conjugate beta-lyase; kynurenine-oxoglutarate transaminase (EC 2.6.1.7),
cytosolic; kynurenine-oxoglutarate transaminase (EC 2.6.1.7), mitochondrial
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 17-Mar-1999
C:Accession: S66270
R:Mahebe, P.; Alberati-Giani, D.; Koehler, C.; Cesura, A.M.
Submitted to the EMBL Data Library, May 1995
A:Description: Identification of a mitochondrial form of kynurenine
aminotransferase/ glutamine transaminase K from rat brain.
A:Reference number: S66270
A:Accession: S66270
A:Molecule type: mRNA
A:Residues: 1-457 <MAL>
A:Cross-references: EMBL:Z49696
R:Mahebe, P.; Alberati-Giani, D.; Koehler, C.; Cesura, A.M.
FEBS Lett. 367, 141-144, 1995
A:Title: Identification of a mitochondrial form of kynurenine
aminotransferase/glutamine transaminase K from rat brain.
A:Reference number: S66269; MUID:95317413
A:Accession: S66269
A:Molecule type: mRNA
A:Residues: 1-71:432-457 <MAN>
A:Cross-references: EMBL:Z49696
R:Perry, S.J.; Schofield, M.A.; MacFarlane, M.; Lock, E.A.; King, L.J.; Gibson,
M.O. Pharmacol. 43, 660-665, 1993
A:Title: Isolation and expression of a cDNA coding for rat kidney cytosolic
cysteine conjugate beta-lyase.
A:Reference number: 157944; MUID:93275300
A:Accession: 157944
A:Molecule type: mRNA
A:Residues: 35-457 <RES>
A:Cross-references: GB:S61960; NID:9300515; PID:9300516
C:Genetics:
A:Genome: nuclear
C:Keywords: alternative initiators; aminotransferase; mitochondrion

F.1-32/Domain: transit peptide (mitochondrion) #status experimental <TNP>
 F.33-457/Product: kynurenine--oxoglutarate transaminase / asparagine--oxo-acid
 transaminase K, mitochondrial #status experimental <MNT1>
 F.33-457/Product: kynurenine--oxoglutarate transaminase / asparagine--oxo-acid
 transaminase K, cytosolic #status experimental <MA12>

566270 Length: 457 February 11, 2000 15:52 Type: P Check: 7215

1 MERSAALSY HLMPLMGK AGASITRCLH QSLTMKRLQ ARRDGIDQN
 51 LWVEGKILK EYDVNLGOG PDESPDPA TQAFQATSG NFMNLQYRA
 101 FGYPPLTNVL ASFFGKLLGQ EMDPLTNLV TVGAYGALF RQALYDGD
 151 EYIMEPAFD CYEPMTMAG GCPVETIKP SPAPKGLGA SMDQDLDAE
 201 LSKFTPRK ILVLTNPNNP LGRVSRMEL ELVANLCOQH DVCISDEVY
 251 QWLVDGHOH VSIALPGMW DRLITISAG KSFATGKWY GWYGPDMIM
 301 KHLRTVHNS IFHCPTAOA AVAOCFEREQ QHFGQPSSTF LQLPQAMEIN
 351 RDMIRSLQS VGLKWLISOG SYFLADISD FSKRPDLPG AEDEPYDRF
 401 AKMINMGL VGIPSTFFS RPHQKDFDHY IRFCFVKDKA TLOAMDERLR
 451 KMKELQ

11AA_SEQUENCE 1.0
 P1:S41286 - latexin - rat
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 10-Sep-1997
 C:Accession: S41286
 R:Hanawa, Y.; Urata, Y.; Takiguchi-Hayashi, K.; Onon, A.; Satoh, K.; Miyamoto, M.; Aizawa, Y.
 Submitted to the EMBL Data Library, December 1993
 A:Description: Intracellular regionally represented specific transcription for a novel protein latexin.
 A:Reference number: S41286
 A:Accession: S41286
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1-223 <MNT>
 A:Cross-references: EMBL:X76985; NID:g440575; PID:g440576

S41286 Length: 223 February 11, 2000 15:52 Type: P Check: 8455

1 MEIIPPIHY ASRAASYAEN CINOQGGPN KVFYQTVQO ASKEDIPERG
 51 HRYHLKFSVE EITKOYTVS CTAELVYPRM GQGSAPENVF TFEGRGNP
 101 DEEDNTEYOR LMSKKEPLQA QNIPDNFGNV SPQKRPVHNL AMVACGYVM
 151 ONSTEDTWIK MAKIQTVKOV ORNDDEFIELD YTVLHDOVAS QELIPMOKOV
 201 LMHPQYGVY KHSRLPKEA PAE

11AA_SEQUENCE 1.0
 P1:A39297 - somatostatin receptor - rat
 N:Alternate names: probable G-protein-coupled receptor; SRIF receptor
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 31-Oct-1997
 C:Accession: A39297; A45102; S20088
 R:Meierhof, W.; Paust, H.J.; Schoenrock, C.; Richter, D.
 DNA Cell Biol. 10, 689-694, 1991
 A:Title: Cloning of a cDNA encoding a novel putative G-protein-coupled receptor expressed in specific rat brain regions.
 A:Reference number: A39297; MUID:92096119
 A:Accession: A39297
 A:Molecule type: mRNA
 A:Residues: 1-391 <MEV>
 A:Cross-references: GB:X61314; GB:X61630; NID:g56309; PID:g56310
 A:Experimental source: brain

A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated 5' to the sequenced region
 R.Li, X.J., Forte, M.; North, R.A.; Ross, C.A.; Snyder, S.H.
 J. Biol. Chem. 267, 21007-21012, 1992

A:Title: Cloning and expression of a rat somatostatin receptor enriched in brain.
 A:Reference number: A45102; MUID:93016064

A:Accession: A45102
 A:Status: Preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-391 <DL1>
 A:Experimental source: brain
 A:Note: Sequence extracted from NCBI backbone (NCBI:P116692)
 C:Keywords: G protein-coupled receptor; glycoprotein; receptor; transmembrane protein

A39297 Length: 391 February 11, 2000 15:52 Type: P Check: 9240

1 MFPNGTASP TSPSSSPG CGEYCSRGF GSGADGMEH PGRNSONGT
 51 LSEGGGSATL ISFTYSVCL VGLGNSMVI YVILRAXKK TATNIYIHL
 101 AINDELMLS VPELVISTLL RHPFGALLC RLVLSDAVN MFTSYCLTV
 151 LSYDRVAVY HPKARFRR PTVAKVNLG VWVLSLVL PIVFSRTAA
 201 NSDGTACNM LMPAQORWL VGFVLTFLM GFLLPVGAIC LCVYLIAKM
 251 RMVALRAGNQ QKRSEKIT LVMYVYVWF VICMPPYV QLVNFAEQD
 301 DATVSLSVI LGVNSCANP ILVGLSDNF KRSPRLICL SMDNAAEFP
 351 VDYATALKS RAYVEDPQ ENLESGVFR NGTCASRIST L

11AA_SEQUENCE 1.0
 P1:JC4294 - ammonia monooxygenase (EC 1.-.-.-) A chain - Nitrosospira sp.
 C:Species: Nitrosospira sp.
 C>Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 21-Mar-1996
 C:Accession: JC4294
 R:Klotz, M.G.; Norton, J.M.
 Gene 153, 159-160, 1995
 A:Title: Sequence of an ammonia monooxygenase subunit A-encoding gene from Nitrosospira sp. NPAV.
 A:Reference number: JC4294; MUID:96001263
 A:Accession: JC4294
 A:Molecule type: DNA
 A:Residues: 1-274 <KLO>
 A:Cross-references: GB:U20644
 A:Experimental source: NPAV
 C:Keywords: monooxygenase; oxidoreductase
 A:gene: amoA

JC4294 Length: 274 February 11, 2000 15:52 Type: P Check: 404

1 MSRTDELKA AKMPSEYKM SMIDAIYFP ILCLILVGY HMFMLLAGD
 51 WDFWDMKDR QWMPYVPIV GITTCATINY DLWVNRLEP GATLICVCL
 101 VGEMLTRFG FYWSHYPIV EYLPSTMIPG ALINDYTMAL TRNMNITLV
 151 GGCARLLRY PGNMPIFGPT HPLVAREGL LSLADYGFEL YVRTGPEVY
 201 RLIEQSLRT FGHRTIYIAG FFSAFYSMLM FCVMWYFGKL YCTAFYVYKG
 251 PGRVITMKND VTAVGEQFP EGIK

11AA_SEQUENCE 1.0
 P1:C75042 - aspartate carbamoyltransferase, catalytic chain (pyrB) PAB1498 -
 Pyrococcus abyssi (strain Otsay)
 C:Species: Pyrococcus abyssi
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: C75042

```

R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Treponema abyssii genome sequence: insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: C75042
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KAW>
A:Cross-references: GB:A248287; GB:AL096836; NID:g5458657; PIDN:CA550232.1;
PID:g516130; PID:g5458745
A:Experimental source: strain Orsay
A:Genetics:
A:Gene: PAB1498

C75042 Length: 308 February 11, 2000 15:52 Type: P Check: 9304 ..

1 MDNRGRDVIS IDFSKEDIE TVLATAERLE RELKEGQLE YAKKILATL
51 FFEPSIRTRL SFESAMHRLG GAVIGFAEAS TSSVKKGEEL RDTIKTYEQY
101 CDVIVIRHPK EGAARLAAY AEVPIVAGD GSNQHPOTL LDYTIKKEF
151 GRIDGLKIGL LGDLKGRFV HSLAEALTFY DVELYLISPE LLRMRPHIVE
201 ELREKGRKV ETTLEDVYG KLDVLYTRI OKERPDEOE YLAKVGSYOV
251 NLKYLEKARD ELRIHNPDR VDEIHPEVDN TKHAIFYROV ENGVPVRMAL
301 LAVLGVY

!!AA_SEQUENCE 1.0
PI:D71333 - conserved hypothetical protein TP0373 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
A:Accession: D71333
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Richardson, D.; Howell, J.K.; Salzberg, S.; Peterson, J.; Khaliq, H.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; McDonald, L.;
Horst, K.; Roberts, K.; Wathney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: D71333
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-477 <COL>
A:Cross-references: GB:AE001216; GB:AE000520; NID:g3322647; PID:g3322651
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0373

D71333 Length: 477 February 11, 2000 15:52 Type: P Check: 4876 ..

1 MSESROKILHP LLYVHARSFG HFLVPRKPSG LLYVNSGGAD SLALTYAHE
51 LADPGVCAC AVTVDSHSLA QEGALDARFV RALCARFSP LPFCVQOISA
101 GAVHACAKIR GRGVODAAARA LRYKVEDHVA ARCGAUVLT AHTRDOYER
151 LLMRLFOGAA ASALQIGRAA RGRVPRPLK VSRTEVEFL QTVGVNRRED
201 ASNTCKRYVR NRIRHELIPA LDAVLAGWS GLDKTFAGIS AHSFSCVAAL
251 TWRREGCSHA WEPVPRALGI RLMPRSDFL AAEFLRFLF LQACVRLGV
301 SHRVPGALE RCARFDGVR IHVSGQLER AGATVLESCI HASDARETK
351 KQDAGSPSS EKQGVSAIV ARGATPCAC GTLIVEVRA GVEVCAODH

401 VGVGPSEFPE YIRTHRTGDR ISIRGNHKG I RMESEWHVP LSDRTVPMI
451 EDDGVLRALY GAALGYONRY AERTPHE

!!AA_SEQUENCE 1.0
PI:G71287 - conserved hypothetical protein TP0730 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-1999
A:Accession: G71287
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson,
R.; Gwin, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.;
Richardson, D.; Howell, J.K.; Salzberg, S.; Peterson, J.; Khaliq, H.;
Artach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; McDonald, L.;
Horst, K.; Roberts, K.; Wathney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis
spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: G71287
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-365 <COL>
A:Cross-references: GB:AE001245; GB:AE000520; NID:g3323033; PID:g3323035
A:Experimental source: strain Nichols
A:Genetics:
A:Gene: TP0730

G71287 Length: 365 February 11, 2000 15:52 Type: P Check: 2410 ..

1 MARASSPLSS WSPHSUPFS AHCAARLRK TCQSGSDASF OYKAHEKST
51 LPAPARRRF PVQYFVSPD KRWILGYFSP PSMKRETAGI LCSYLFHSA
101 LVLAISLHAG RTOVPPSSLT SELVIVLMH CLLEFVLAV SNEPDTIVP
151 FXPDEQAPR ICAAAASDCK ENRTALKTIN TATHILINA SAIPVIGTL
201 KHALAGLSY FLVAGLSVLF LTDFIDGRIA RARETSRNG ETLDAASDYA
251 LIGLISALY QSGVPLMEF VLIITRSLQ TVIACYALF GHPMTGSTAG
301 GKATVAVTML LYTLELRL LPNLARSNSG ARRFAGETL AGVITFTGIV
351 EKLYGVQHR PGRSP

!!AA_SEQUENCE 1.0
PI:J00857 - hypothetical 18.1K protein - Escherichia coli retron Ec67
N:Alternate names: regulatory protein CII homolog
C:Species: Escherichia coli retron Ec67
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 10-Sep-1997
A:Accession: J00857
R:Hsu, M.Y.; Inouye, M.; Inouye, S.
Proc. Natl. Acad. Sci. U.S.A. 87, 9454-9458, 1990
A:Title: Retron for the 67-base multicopy single-stranded DNA from Escherichia
coli: a potential transposable element encoding both reverse transcriptase and
Dam methylase functions.
A:Reference number: J00851; MUID:91067724
A:Accession: J00857
A:Molecule type: DNA
A:Residues: 1-169 <HSD>
A:Cross-references: GB:M55249; NID:g145143; PID:g145147
A:Experimental source: E. coli strain Cl-1
A:Note: Insertion site is equivalent to 19 min of E. coli K12 genetic map

J00857 Length: 169 February 11, 2000 15:52 Type: P Check: 6140 ..

1 MFDYOVSKHP HEDACRARA IRRNLVOLAE RACMNOIIR NKLNSQPHL
51 LTAPEIWLIT DLEDSTLYD GFLAQIHCLP CVQINEVARE KLPHVMSAT
101 AEIGVAAAGA VSGDVYTSAG RDAISSINS VTRMLALAV SLQARLQANP

```

151 AMASAVDTVT GIGASGILL

!!AA_SEQUENCE 1.0
 P1:T10288 - hypothetical protein 19 - Orygia pseudotsugata nuclear polyhedrosis virus
 C:Species: Orygia pseudotsugata nuclear polyhedrosis virus, OPMNV
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
 C:Accession: T10288
 R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.
 A:Title: The sequence of the Orygia pseudotsugata multinucléocapsid nuclear polyhedrosis virus genome.
 A:Reference number: 217011
 A:Accession: T10288
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-298 <AHR>
 A:Cross-references: EMBL:U75930; NID:92934903; PID:g1911265
 T10288 Length: 298 February 11, 2000 15:52 Type: P Check: 562

- 1 MLVQVNFLO LVLAHALEFL CSFAEYFALM ATVATARAFL LELEDSABSI
- 51 INSHLAFL LGPYLATIT WAMTKMLICY KGLEKRSNFY MKTVVLAHL
- 101 MAGSCMLFV VFQPIHNG HVPYLDALR HHDRSLCWS GYVVOEYEH
- 151 DANIRFDLN CVYDNFMK CVGCRMVHR DEPTVFQNO GALTMLALLA
- 201 IYMCWNNVY OQKERRKPN RARNITLTL METEKYDPA EEEHESNMNR
- 251 SMISARRAR NRPHNYFPC RQSGOTONS RLIFQKRAA ROVPTAE

!!AA_SEQUENCE 1.0
 P1:G72757 - hypothetical protein APE0052 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: G72757
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Funahashi, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.; Aoki, K.; 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
 A:Reference number: A72450; MUID:99310339
 A:Accession: G72757
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-175 <RAW>
 A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA78961.1; PID:d1042737;
 A:Experimental source: strain K1
 A:Gene: APE0052
 G72757 Length: 175 February 11, 2000 15:52 Type: P Check: 9964

- 1 MSPALIRITA IAAVLAVFI AIGAVAIYTS EYSDVSTLOS VTRASRTVQ
 - 51 AGVAVIGYGT ATVIYGKTY TLEARGAYCI LMPIDGSGS YAFVNEGEK
 - 101 GYVAALYEL DSFTARRGGS PYEDIVVVD GYVRGEELV LTPAGEEST
 - 151 PVVTNAILK GCHAIYSEK AVEVO
- !!AA_SEQUENCE 1.0
 P1:C72759 - hypothetical protein APE0064 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C:Accession: C72759
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Funahashi, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.; Aoki, K.; 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
 A:Reference number: A72450; MUID:99310339
 A:Accession: C72759
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-114 <RAW>
 A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA78973.1; PID:d1042749;
 A:Experimental source: strain K1
 A:Gene: APE0064
 C72759 Length: 114 February 11, 2000 15:52 Type: P Check: 3453

- 1 MTSLSAGSIA SPLWYTRST AAPARARA ASSIRASSG VRCVSSRS
- 51 TALEPRARAL ITSSLISEN PISFKYLSGS INPRASDAS ASIYACLRP
- 101 AFLPATNRVT ARSS

!!AA_SEQUENCE 1.0
 P1:A72773 - hypothetical protein APE0172 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: A72773
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Funahashi, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.; Aoki, K.; 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.
 A:Reference number: A72450; MUID:99310339
 A:Accession: A72773
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-129 <RAW>
 A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA79083.1; PID:d1042859;
 A:Experimental source: strain K1
 A:Gene: APE0172
 A72773 Length: 129 February 11, 2000 15:52 Type: P Check: 7998

- 1 MSPLISIR GADARVSTS LFSTGFANR VTCICMNV GTATRAVMS
- 51 SGGSTVLTST AMWIGRGGL RTIESIGPEL TSLGAPLIT RTALTISTPN
- 101 SLFRVHISLR RPYVLDHTR RLPLPLSS

!!AA_SEQUENCE 1.0
 P1:C72775 - hypothetical protein APE0189 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: C72775
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.; Funahashi, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.; Aoki, K.; 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.

Mon Feb 14 08:07:32 2000

pir.cat

Page 182

A:Reference number: A72450; MUID:99310339
A:Accession: C72775
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-100 <RAW>
A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BA79101.1; PID:d1042877;
A:Experimental source: strain K1
A:Gene: APE0189

C72775 Length: 100 February 11, 2000 15:52 Type: P Check: 1302 ..

1 MLEFRTSD IIFDLKLAP VIIIPMSIS VSRSPILCS TLVIMKAM

51 ASSSVASCFR DILACAVSR MRASCLVYM GTGFAPSMLE WRRALLIST

!!AA_SEQUENCE 1.0

P1:C72784

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999

C:Accession: C72784

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Nagai, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic

Crenarchaeon, Aeropyrum pernix K1.

A:Reference number: A72450; MUID:99310339

A:Accession: C72784

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-313 <RAW>

A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BA79173.1; PID:d1042949;

A:Experimental source: strain K1

A:Gene: APE0258

C72784 Length: 313 February 11, 2000 15:52 Type: P Check: 9325 ..

1 MMLVGGSLDD VKEVSRLG DKDFQGLMK SGLNOKKVA ELLSHPEIL

51 EAARKYTEVK RESYKRLGEL VERAVESLR VGANPIYAEI AEDAREIVK

101 IYGGSIIVYM SKSWAEIG LREHESMGN EWEETDLOI LVMLEGVKPM

151 HNVAVAVMT RERARIRE KLGIEVDERD VEGNVAAYRE FLREKFFKAR

201 VGITGANSMS ADTGITIVLE NEGNIRLVSS LPPVVAALVP IDKIVPSVMD

251 AVNVALVOAA FAGFMPVYI SLITGPSATG DIEQIKVLGA HGPREVAVVL

301 LDNGMKKAS HPE

!!AA_SEQUENCE 1.0

P1:H7271

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999

C:Accession: H72721

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Nagai, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic

Crenarchaeon, Aeropyrum pernix K1.

A:Reference number: A72450; MUID:99310339

A:Accession: H72721

A>Status: preliminary

A:Experimental source: strain K1

A:Gene: APE0339

H72721 Length: 168 February 11, 2000 15:52 Type: P Check: 4962 ..

1 MMSGVEIVLR PIGVSHGLP DSTVRENIGG VEGFIEVPE YAGGLDGLAG

51 FSHIIVAYL HSKGRALRV RPRGLRLGF RLEDLEPGV FATDPSRPN

101 PLALTIARLL DIEGRRLRVS GLDLYDGTVP LDKPIYPSR RVEDLSLPSM

151 YKELLEASR RDRVRREV

!!AA_SEQUENCE 1.0

P1:H72731

C:Species: Aeropyrum pernix

C>Date: 20-Aug-1999

C:Accession: H72731

R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, F.; Nagai, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic

Crenarchaeon, Aeropyrum pernix K1.

A:Reference number: A72450; MUID:99310339

A:Accession: G72736

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-160 <RAW>

A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BA79387.1; PID:d1043173;

A:Experimental source: strain K1

A:Gene: APE0393

H72731 Length: 172 February 11, 2000 15:52 Type: P Check: 2984 ..

1 MGSILSSSI PISSSVYLA SSAALAIIPS VMADLTTRA PLKTPGNTRA

51 LTLWGRSL PVATLAPAS LASHGHISGV GFASANTMGS LDMPLHSFL

101 TTPGLEANA TRASAPLRAS SKLLPGASPP PSVSRASLYL YLEKRSLSI

151 SSRSGWRPL LSTTMTLGL TP

Mon Feb 14 08:07:32 2000

plc.cat

Page 183

```
C:Genetics:
A:Gene: APE0429
G72736 Length: 180 February 11, 2000 15:52 Type: P Check: 2336
1 MSASVARRRR ESKKEGEGE KSKDKLIISV EWTNRDVKV NPMRALIYL
51 IDRLGPIHER TLQITAELO QRGALGYEF KIVAGVPSP EFKNDLIALA
101 YGVEVEVNR RNRLQTTND GKALEKHGA PKGVEVELEK HYDEINIAS
151 LEDKVDDEL KITKOLKGR RREFGGLGF

!!AA_SEQUENCE 1.0
PI:D72737 - hypothetical protein APE0434 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72737
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <KAW>
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BAW992.1; PID:dl043178;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0434
D72737 Length: 120 February 11, 2000 15:52 Type: P Check: 7680
1 MRSPFMSY LYFINLSE NLGCGGTATG PPTGASIR SVYPLHLHF
51 LRSHONLSP QPPLSLRR ILRPVKKSS WLQPSSEPF SPASCSPPA
101 APLCTWGRRR CISMILLSL

!!AA_SEQUENCE 1.0
PI:A72741 - hypothetical protein APE0458 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72741
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: A72741
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <KAW>
A:Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BAW9421.1; PID:dl043207;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0458
A72741 Length: 160 February 11, 2000 15:52 Type: P Check: 338
1 MWSGAILTA ETPAPITRR LKLDISTOA RASIMGSIYF SLSITSTPII

!!AA_SEQUENCE 1.0
PI:G72640 - hypothetical protein APE0559 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72640
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: G72640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <KAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAW9527.1; PID:dl043313;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0559
G72640 Length: 147 February 11, 2000 15:52 Type: P Check: 9276
1 MWTTPYPRAA STASTRPGS HGASMTPTPL CLAMTSISLT ALAGSPSSSS
51 WSLGTPEEP PILTPLETF HSTAAAPAR ACLQASTEP LDEKTPPT
101 SSSGSPPEP AIRRAAPST SSAALATRA RVLETGSLAG ADKLLHP

!!AA_SEQUENCE 1.0
PI:A72651 - hypothetical protein APE0638 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72651
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Anka, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: A72651
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-119 <KAW>
A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAW9609.1; PID:dl043395;
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE0638
A72651 Length: 119 February 11, 2000 15:52 Type: P Check: 639
1 MTRLRDPPA GAGGIYSRC LSNTPFSTLA GFRASARTSK ALMGFWPPO
51 LYSTIVANLY SGHWYTWML SRIMATTLTP SPIODIMSR TVMPSPRAAS
101 RTTSSSSGL FRSSVLAHS

!!AA_SEQUENCE 1.0
PI:D72672 - hypothetical protein APE0802 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
```

```

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: D72672
R: Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A/Reference number: A72450; MUID:99310339
A/Accession: D72672
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-129 <KAW>
A/Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA79780.1; PID:d1043566;
PID:95104465
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0802

D72672 Length: 129 February 11, 2000 15:52 Type: P Check: 4161 ..

1 MYPIKSLPS QLSCLINSL EMSNFTSRSE PSSPISPENL GTLLHGLVTL
51 NPFLGLGMLT TINFALLRV STDVSTARE MLRASGROSV SSSIQAELIE
101 TDLNPLRAAS RAPKVPLEK KSYTPRLGP

11AA_SEQUENCE 1.0
P1:E72679 - hypothetical protein APE0857 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: E72679
R: Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A/Reference number: A72450; MUID:99310339
A/Accession: E72679
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-234 <KAW>
A/Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA79837.1; PID:d1043623;
PID:95104522
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0857

E72679 Length: 234 February 11, 2000 15:52 Type: P Check: 8681 ..

1 MLNRKNQO VEYCGPRGG QDELEGVYH PPTVAPRQL AGAEMDLYE
51 PVDAGAEIV EMHAGYPLVL PVPAPRQAR ILIRYAVHGL ARRRVYHRS
101 QPLDVELTLD VGAVDLGVAG GGGRMHLEE RRODCSEDE SRSYLITRP
151 PKRLCVNRRR PEERKNGEG VEAVDSLCPA SRGCLSLIGA VDRLPORAAG
201 RVYGHRRPP GSRVALLPHY NGPGVAPVIO RGGG

11AA_SEQUENCE 1.0
P1:F72685 - hypothetical protein APE0902 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: F72685
R: Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A/Reference number: A72450; MUID:99310339
A/Accession: F72685
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-228 <KAW>
A/Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA79886.1; PID:d1043672;
PID:95104571
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE0902

F72685 Length: 228 February 11, 2000 15:52 Type: P Check: 9005 ..

1 MEALYRVEQ LVLNMGIDY LGLSDEIRD VLEPIVLEAA SKLSRPAEC
51 IARKIIQKH LFLKALASRL AESYDRUTRE RLEPIVONAP EIGKKAAPAL
101 YDVARRIGAD DIVEELRLM EAVGSPTKAR CPKCGFKALT PELMCLVCKR
151 TSSEEFKEX IGFOGLERW ASRAPEIVE EYLRSSIIY EDGALAAPSE
201 PKTPLAVPLT LGSREKSIYK KRLERLR

11AA_SEQUENCE 1.0
P1:H72607 - hypothetical protein APE1326 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix
C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C/Accession: H72607
R: Kawabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A/Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A/Reference number: A72450; MUID:99310339
A/Accession: H72607
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-387 <KAW>
A/Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80318.1; PID:d1044104;
PID:95105004
A/Experimental source: strain K1
C/Genetics:
A/Gene: APE1326

H72607 Length: 387 February 11, 2000 15:52 Type: P Check: 1372 ..

1 MAYSNQYS LVDAVHIKRL VAEELRRYV EEGFRASRD WHRRRPIPC
51 GMTHIGVCG SYGCLYCIY DMGFTGKPOD YPLSDELAL ALALNPYAP
101 GPKGTLIAFG SYTEPFMRK AMRAIEYLS TRWIGNPOQ ISTKIALRG
151 LLERFIEMAD PRIDVLVMT TLSRMWALP GASPEERIE FMGRIVEAGL
201 SATLELPII PGVTDREAED ILSRAARAG NKVVLGTLRV AEGILRLRA
251 SGAVGEKEI RRLPWRPKG EQLPIYSRL KERLSKARE MGRKVLPAAC
301 SANVESGQG CAACRLPGC DLSKPSAGE REVALLEAL GLPVPKYSV
351 GHSVEAVLNG GRREAISYI WIIGLLRRYP RVRTIVRN

11AA_SEQUENCE 1.0
P1:E72626 - hypothetical protein APE1469 - Aeropyrum pernix (strain K1)
C/Species: Aeropyrum pernix

```

```

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E72626
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72626
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KAM>
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80467.1; PID:d1044253;
PID:95105153
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1469

E72626 Length: 146 February 11, 2000 15:52 Type: P Check: 6841 ..

1 MSTVMDRTE ISLVPRFTM RGTPAYLRL ATRALAVALD RLVTGMSHL
51 MLEQMRQWV RELQSEASC NAEPSGLNL METYAKALPG VMLAAVPALS
101 LVTVALLSRS RMARALISM SAAPLSGLEE WSPMPSHSIS IYRLPL

11AA_SEQUENCE 1.0
P1:D72630 - hypothetical protein APE1499 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72630
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72630
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <KAM>
A:Cross-references: DDBJ:AP000061; NID:95104821; PIDN:BA80499.1; PID:d1044284;
PID:95105184
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1499

D72630 Length: 152 February 11, 2000 15:52 Type: P Check: 4651 ..

1 MSLRSRSRKR RSMRGTSST IYTLAIPGLI LGWSLSCIMS TLHLFSAIT
51 SRTIYSWPMG SCITPVILAI LSLARPTS I TSMVLVSAP PLITITRVL
101 LGTTIARRA ATPAPAPSA RIIDLSRRRY MALAISCSST ATTLSTRCT
151 MG

11AA_SEQUENCE 1.0
P1:H72538 - hypothetical protein APE1597 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H72538
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72537
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-175 <KAM>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BA80597.1; PID:d1044383;
PID:95105284
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1597

H72538 Length: 175 February 11, 2000 15:52 Type: P Check: 9362 ..

1 MDECHRLVGY IYIYLANMY RVLGSRLLHA HRHRDRKDV NPVPACRTV
51 VMLGAGTGD GEVEYLLIGY VASVYVNPV VGRYRGQI QYQALRLLD
101 RRYVAHNSNL AADDRHGGE CYGLPVPHPH NLHLDPGRQ PLGLPAVED
151 PLSLENKPLQ GGERGVAGV AGVVG

11AA_SEQUENCE 1.0
P1:D72545 - hypothetical protein APE1648 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72545
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: D72545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-194 <KAM>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BA80649.1; PID:d1044435;
PID:95105336
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1648

D72545 Length: 194 February 11, 2000 15:52 Type: P Check: 2785 ..

1 MWGRGLKQWT LGDRVLAISM GDLTKVRAEA VYVNPANSLMI MGGAGAGLK
51 RAGGSVIEEE AMRKAPYPVG EAVITSGSL PARFVIHPT MEEPMRIFL
101 VNAFRASYAA IRLASEAGIE SVAMPAGAG VGLSLVAEVA REAMPAASIL
151 RGMWPRTIIL VARGEAYRG MEKGVREALG VEGGECPADL ARLV

11AA_SEQUENCE 1.0
P1:E72557 - hypothetical protein APE1745 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E72557
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankai, A.; Kosugi, H.; Hosoyama, A.;
Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1.
A:Reference number: A72450; MUID:99310339
A:Accession: E72557

```


Mon Feb 14 08:07:32 2000

pir.cat

Page 186

```
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1132 <KAW>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80746.1; PID:d1044532;
A:Experimental source: strain K1
A:Gene: APE1745
E72557 Length: 132 February 11, 2000 15:52 Type: P Check: 8347
1 MRVLPRGFS LTASSRYMAL EPESILITSA CRINERAAVY TFFSNAPAPR
51 TLPGTMTISP SLACFELILR LTSALLIRGL SRSATLSQT GAFELALALL
101 SSPISLRVTR LVGPVGLMTP LHPHSLRPLQ TS
!!AA_SEQUENCE 1.0
P1:H72557 - hypothetical protein APE1748 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
Fukui, S.; Nagai, Y.; Sekine, M.; Baba, S.; Aoki, A.; Kosugi, H.; Hosoyama, A.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, U.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: H72557
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1181 <KAW>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80749.1; PID:d1044535;
A:Experimental source: strain K1
A:Gene: APE1748
H72557 Length: 181 February 11, 2000 15:52 Type: P Check: 8142
1 MFGSSVLLG SPTSMVLSCS KRAVRSLLI LSIILSSSS LSHRNKSTR
51 PRAIILIAS AMPITPPPT LTLMSTRLLT PSPAIRGSI IFILAIISGY
101 ISRGPIPIQT LPLPLITAL AIAFLLEVE ITILSTAIPC PLTNSQHPSS
151 TRAAQRHOA LCSESIAMRH RPRGTSRTPC L
!!AA_SEQUENCE 1.0
P1:D72577 - hypothetical protein APE1900 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
Fukui, S.; Nagai, Y.; Sekine, M.; Baba, S.; Aoki, A.; Kosugi, H.; Hosoyama, A.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, U.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: D72577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1100 <KAW>
A:Cross-references: DDBJ:AP000062; NID:95105244; PIDN:BAA80905.1; PID:d1044691;
A:Experimental source: strain K1
A:Gene: APE1900
D72577 Length: 100 February 11, 2000 15:52 Type: P Check: 2318
1 MIPVVFINT SMRSEFGKSO TAMEPVLRSS ILSGHINSM DILITGSSIP
51 SSIISFNFS LKLSATSMSEI EEIKWPISA RTATRPPIPR ADPIVETPT
!!AA_SEQUENCE 1.0
P1:A72511 - hypothetical protein APE2063 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
Fukui, S.; Nagai, Y.; Sekine, M.; Baba, S.; Aoki, A.; Kosugi, H.; Hosoyama, A.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, U.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: A72511
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1140 <KAW>
A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA81073.1; PID:d1044859;
A:Experimental source: strain K1
A:Gene: APE2063
A72511 Length: 140 February 11, 2000 15:52 Type: P Check: 2835
1 MLSASSAMS DMKRTILATA SRASILNLSLT LPISLNLSPK AFSNTSAGP
51 LNSMLWYITF SHSPSSILATI PALTRAGVWS SASSSGLPFS LHSLLQSNII
101 LHSKAGAMPS LAPISQLRFQ LNTSRPAMSI VYNATASSSES
!!AA_SEQUENCE 1.0
P1:C72529 - hypothetical protein APE2207 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
Fukui, S.; Nagai, Y.; Sekine, M.; Baba, S.; Aoki, A.; Kosugi, H.; Hosoyama, A.;
Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, U.; Kushiida, N.; Oguchi, A.;
Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
Crenarchaeon, Aeropyrum pernix K1
A:Reference number: A72450; MUID:99310339
A:Accession: C72529
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1310 <KAW>
A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAA81219.1; PID:d1045005;
A:Experimental source: strain K1
A:Gene: APE2207
C72529 Length: 310 February 11, 2000 15:52 Type: P Check: 9995
1 MRAEVRVTF SDIVDIYAYN YFSWLAVALV KPELEBRNLK IAGIQVYFVM
51 YNSRLILIST AVFLGCVFS FWTLISGSI IYKAIYVLLA LSAFPMFPA
101 GLMPSLKTG SRKEGIETEL PFFAAYLSIM GRCGVVSVLY IDRYASLKIT
151 KALREAEEMI RTKIRILGKN PLDALEHVL DSSSVFNF ILGTTAVKY
```

Mon Feb 14 08:07:32 2000

pir.cat

Page 187

```
201 GSDVHLEI RTODLEFRL NDIRLAERN TLLEYIIV AVIATIVERI
251 FTISSEVHC NLGVASAOY ASTTOLINS FLVLEPIVI MLVDRARP
301 KEPIQIDTW

11AA_SEQUENCE 1.0
PI:D70471 - hypothetical protein APE2415 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: F72471
R:Accessions: Y: Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, J.; Baba, S.; Anket, A.; Kotsugi, H.; Hosoyama, A.; Fukui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takemida, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Oguchi, A.; Funahashi, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
AOKI, K. 6, 83-101, 1999
DNA Ref. Complete genome sequence of an aerobic hyper-thermophilic
A:Title: Complete genome sequence of an aerobic hyper-thermophilic
A:Reference number: A72450; MUID:99310339
A:Accession: F72471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-113 <RAM>
A:Cross-references: DDBJ:AP000064; NID:95105945; PIDN:BA81430.1; PID:di045216;
PID:95106119
A:Experimental source: strain K1
A:Genetics:
A:Gene: APE2415

F72471 Length: 133 February 11, 2000 15:52 Type: P Check: 430 ..

1 MLVKRPTSL LPTIATAPI ESRPTTRAP VYALTALEM LLARFPTPT
51 VAVAMIKPKV ADWTMCINTS DILYATIGA PSLATIPIMT RNPMLITSIM
101 RETGPTRSI CPISCRSGL GFLIKPGAPL RSR

11AA_SEQUENCE 1.0
PI:C70405 - hypothetical protein aq_1220 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 08-May-1998
C:Accession: C70405
R:Accessions: G: Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, R.; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Huber, R.; Feldman, R.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Swanson, R.V.
R.A.: Short, J.M.; Olson, G.J.; Swanson, R.V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: C70405
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-499 <AOF>
A:Cross-references: GB:AE000729; NID:92983659; PID:92983669; GB:AE000657
A:Experimental source: strain VFS
A:Genetics:
A:Gene: aq_1220

C70405 Length: 499 February 11, 2000 15:52 Type: P Check: 5299 ..

1 MMRKFLVLI PVLLIYFNL GYNNAMPNE SFYDSAKNM LKSGFLTPV
51 YNGEFLNKP PMTWIVSLG YKIVGNELG LRFHALGV FTGVLTYLLA
101 RRTGSKNTA LLSFLILIS FIFTANARY SPEVPTFFI TLSLYMYEX
151 FTRKRELFW LALIASLAV LTKGPAGVL PAGVFEYLL LRAPEKLLKI
201 KYAGILWVF LLSGMWELQ YIVAREELE VFKEKVKRI YALORDFEYF
251 YALDIVSFL PYSLFFFL FALKERRE LSPFLWFSF IFLIFSIVKM

11AA_SEQUENCE 1.0
PI:D70454 - hypothetical protein aq_1793 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1999 #sequence_revision 08-May-1998 #text_change 08-May-1998
C:Accession: D70454
R:Accessions: G: Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, R.; Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Huber, R.; Feldman, R.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Swanson, R.V.
R.A.: Short, J.M.; Olson, G.J.; Swanson, R.V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: D70454
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-269 <AOF>
A:Cross-references: GB:AE000756; NID:92984076; PID:92984086; GB:AE000657
A:Experimental source: strain VFS
A:Genetics:
A:Gene: aq_1793

D70454 Length: 269 February 11, 2000 15:52 Type: P Check: 3320 ..

1 MMSPMASIM YTHFRMKHA LNWSTILFQ NPLTAEILI LFTLLNELF
51 IPYANVFAEF IHNVNFSLI IYFSKLYIKV KGNBEYKRE IERTKLAQAL
101 KYLPRAVTL TEATYAMTA YILLFVSL ILCLIGITIV FTFGIDPIIA
151 YLIFIVILLI LYFWITSTYP VEFARTVIEG QIPRDEPFLF IIAFESKILW
201 KLFESLEVLF SSVIGFESL FIFLFOFVMS HLFPPLFFLF YVAFSNTLL
251 IYLFGVISVS YLMKRECK

11AA_SEQUENCE 1.0
PI:PA0052 - hypothetical protein PA0052 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A75196
R:Accessions: G: Anonymus; Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: A75196
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-147 <KAB>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CA849016.1;
PID:el514910; PID:95457525
A:Experimental source: strain Orsay
A:Genetics:
A:Gene: PA0052

A75196 Length: 147 February 11, 2000 15:52 Type: P Check: 70 ..

1 MSSLSLAMP FLNSISGAT TLTSFIVAYF KAEPILSTNS LYVSKVSPS
51 SLAKPTNSP PRASKREA TPISKFLAG IOTLLANSLA ESKSSSSSLAV
101 TSGTLPSTNL LPAIFNSCR KSIFTIKGT FMSLEKNSAT LFTYLS
```

```

P1:E75204 - hypothetical protein PAB0107 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: E75204
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: E75204
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1179 <KAM>
A:Cross-references: GB:A1248283; GB:A1096836; NID:g5457433; PIDN:CAB49084.1;
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0107

E75204 Length: 179 February 11, 2000 15:52 Type: P Check: 8207 ..

1 MGCYFNKATN IKIRHIMKR DRGDFLSLS NPEISPIILL VIGSIFPDL
51 VTFIFSFEAL AHGGEIDIRK EHRSTYHSL FLAPLVVSL AKSLFMFTI
101 GAASHLFLDF FSGVIFFPY LRRKGIVKI ILSIGTGST KAKILRRPD
151 PRIERKIEIS KSIYLLTL ILFLACKC

!!NA_SEQUENCE 1.0
P1:H7185 - hypothetical protein PAB2346 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H7185
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution.
A:Reference number: A75001
A:Accession: H7185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <KAM>
A:Cross-references: GB:A1248283; GB:A1096836; NID:g5457433; PIDN:CAB48935.1;
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2346

H7185 Length: 266 February 11, 2000 15:52 Type: P Check: 6767 ..

1 MIRVATFEDI DDMVSFIDA VNYTGRDSV AKSMELSLK QNGCLIAFI
51 DCKPYMGCI FLTKKAWMG LMVKKKQVR RGIETELFR LKIKKGTI
101 RDDASQGYG LYKKNFTDE YRTVRELRN RPLKVSQV ELKEIPGWE
151 ELDKRAFQGD RTVRLIYLE SGARLIVEN EGSGMYRGR IDPLVAISRK
201 IAEKIMLKAF LLRGREIIV DANEDAMLI KKSPEVLTS CYRMLGDKV
251 EENVHVFEG LTYAKG

!!NA_SEQUENCE 1.0
P1:D71228 - hypothetical protein PH0086 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: D71228
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: D71228
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-135 <KAM>
A:Cross-references: GB:AP000001; NID:g3236128; PID:d1030098; PID:g3256472
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by
GenBank
C:Genetics:
A:Gene: PH0086

D71228 Length: 135 February 11, 2000 15:52 Type: P Check: 1741 ..

1 MRGATTLTLE IVAYSKALLT LSTNSFVLSG KVSPPSLNRP TIISPPRAS
51 RREAILIKT KPLAGTHILL AVSLALSTS NSLAETLGRF PSTNALPESL
101 ENSLRKLLI IFKGLISFE KNSATLFTLL SLLP

!!NA_SEQUENCE 1.0
P1:D7110 - hypothetical protein PH0651 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: D7110
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: D7110
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-102 <KAM>
A:Cross-references: GB:AP000003; NID:g3236130; PID:d1030685; PID:g3257059
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by
GenBank
C:Genetics:
A:Gene: PH0651

D7110 Length: 102 February 11, 2000 15:52 Type: P Check: 6672 ..

1 MLTISLGM NIAKENVVY LPEVPATR MFAGLPRPS TRSQNATS
51 EFKVNFIS IIVSGSLNF LIVIGPSRL IGSVALTILL PSGSLASNG
101 FC

!!NA_SEQUENCE 1.0
P1:E71120 - hypothetical protein PH0732 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: E71120
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikku, Y.; Funahashi, T.;
Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya,
H.; Kikuchi, H.
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a
hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137

```

Mon Feb 14 08:07:32 2000

pir.cat

Page 189

A:Accession: E71120
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-104 <KAW>
A:Cross-references: GB:AP000003; NID:93236130; PID:dl030766; PID:93257140
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0732

E71120 Length: 104 February 11, 2000 15:52 Type: P Check: 8797

1 MAHFNHYYF PRLHCHYAI PINEGMHYI VRSSPRDLSC YQFISIRYVS
51 KAKISYLLNS CLLKSTIYI HSTSKNEMPV NLYHRNYILN TFSYNGSFS
101 FINI

!!AA_SEQUENCE 1.0
P1:H71124 - hypothetical protein PH0767 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
A:Accession: H71124
A:Cross-references: GB:AP000003; NID:93236130; PID:dl030801; PID:93257175
A:Residues: 1-272 <KAW>
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Cross-references: GB:AP000003; NID:93236130; PID:dl030801; PID:93257175
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0767

H71124 Length: 272 February 11, 2000 15:52 Type: P Check: 5357

1 MANFNRLSS SSARTRPAP ALVIMATLP LGKGCENALA MSLSSIVFA
51 TKTPCLINAA SYIPOLPAKE PVMLEDAALP LSLPFAFKAM IGFEVTLFA
101 TSMNFFPFLT PSMTYRAIALS SSVSAKYSIS SAKTSAMLP TLNENFPMPS
151 SEIVLYTIPA PNAPLIMLNA TPEPLATVIK VAFKSTWFI IPTQFPMIL
201 IPIYMAISTS SFSLEFPSP VSPNPAVITR TPLTPSPPHW ISVSFTNLFG
251 TTTIAKSSSS GISTLLVLA TP

!!AA_SEQUENCE 1.0
P1:F71127 - hypothetical protein PH0787 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
A:Accession: F71127
A:Cross-references: GB:AP000003; NID:93236130; PID:dl030823; PID:93257197
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0787

A:Reference number: A71000; MUID:98344137
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-183 <KAW>
A:Cross-references: GB:AP000003; NID:93236130; PID:dl030823; PID:93257197
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0787

F71127 Length: 183 February 11, 2000 15:52 Type: P Check: 8012

1 MLILHSEGA SVDCIRTLI RMSKNRLPQ GKITSKIKI STGAFSLVNL
51 ALIVDLAPY KPGIAVEYSV GSKRDLALD LQELKNSVT PELVEPFOI
101 EYTPVTRR TYAIGLVYN KPRKANTKF MIONRRIILA KYLELLNYNI
151 KALNISELAR MEGVSRDITV NDIOQIINKV DKY

!!AA_SEQUENCE 1.0
P1:F71039 - hypothetical protein PH1606 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
A:Accession: F71039
A:Cross-references: GB:AP000006; NID:93236133; PID:dl031661; PID:93258035
A:Residues: 1-447 <KAW>
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Cross-references: GB:AP000006; NID:93236133; PID:dl031661; PID:93258035
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1606

F71039 Length: 447 February 11, 2000 15:52 Type: P Check: 7274

1 MKKIKSGII ISALYTICML YLTQYINGL RYNSDLGIF AQSHSLSGG
51 KLFYNTVERQ LYNVQIHFGV HPQIFLELF PLFLFKKTT YALVVSQITA
101 LGTSVYLAIR LATELDEKR GVALTILYAC NSSLIGINIF EHPVSLAVP
151 LFLAVKFKL RGSPLFYIIS TLLILTKEDA FLVVISIIM KILKGFSEVE
201 ILKRNKRIIF EAPLALVGI LITKYIIRKF GGRYISYIS TRFSIDRRKL
251 LYELLENTLF ALPMLDFIG MSLPFWME CLASRBTQI MGFHYPMIL
301 VPLSFVASIE VARQIKRIY KRVITIGILA SLATLPIAKL PEKPNPLIY
351 LVTITPVGK EASWEAIEIV KGLKPIYTO PEFVPLAIR SDVYIYFKNV
401 NRVILVNLN TYGRRALKR LRDFKVELR YKKIEKGV IIMLRD

!!AA_SEQUENCE 1.0
P1:F71201 - hypothetical protein PH1883 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
A:Accession: F71201

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, G.; Sekine, M.; Babo, S.; Kosugi, H.; Hosoyama, A.; Nagai, T.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikubo, T.; Funahashi, T.; Yoshida, T.; Kudoh, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
 A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.
 A:Reference number: A71000; M01D:98344137
 A:Accession: F71201
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1137 <KAM>
 A:Cross-references: GB:AP000007; NID:93236134; PID:d1031948; PID:93258322
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by C:Genetics:
 A:Gene: PH1996

F71201 Length: 137 February 11, 2000 15:52 Type: P Check: 6730 ..
 1 MSAVITLGL SVSONSESIP CGAVSNMPP PLFLMSKNQL VSIPSKNIGG
 51 FTYSEIYMY SSPISPLSI ISLAFCSGVY NLGKGTISL TPAFLAFLTI
 101 SLPSSVSVM GFSRSISFSS SMALITSLC LYVSIM

!!AA_SEQUENCE 1.0
 P1:D71216 - hypothetical protein PH1996 - *Pyrococcus horikoshii*
 C:Species: *Pyrococcus horikoshii*
 C>Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text-change 14-Aug-1998
 C:Accession: D71216
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, G.; Sekine, M.; Babo, S.; Kosugi, H.; Hosoyama, A.; Nagai, T.; Sakai, M.; Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamiya, M.; Ohikubo, Y.; Funahashi, T.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
 A:Title: Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.
 A:Reference number: A71000; M01D:98344137
 A:Accession: D71216
 A:Status: preliminary
 A:Molecule type: nucleic acid sequence not shown; translation not shown
 A:Residues: 1-225 <KAM>
 A:Cross-references: GB:AP000007; NID:93236134; PID:d1032066; PID:93258440
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by C:Genetics:
 A:Gene: PH1996

D71216 Length: 225 February 11, 2000 15:52 Type: P Check: 2646 ..
 1 MCSFLMIAF DPGTLVDY SCIEAFYFA LRRTYFPFG KRYMKLLTK
 51 IELQFERPRF GRHSGRIKAP MKIFQGFAR IWEERAKLT KELEGAREVL
 101 KELKREGHIV ISPSADFTF GYKERYLQAS GLYDIEDVVI IFGHVSATKE
 151 AFRIVREKYG NETFWVDK PWRFGNGDE NTEYVMATPP YTAPEVTKRI
 201 LDKIPHLHVI YDLMSLDVL RRLSY

!!AA_SEQUENCE 1.0
 P1:G75035 - iron (iii) abc transporter, permease protein (hemu-2) PAB1535 - *Pyrococcus abyssi* (strain Orsay)
 C:Species: *Pyrococcus abyssi*
 C>Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 20-Aug-1999
 C:Accession: G75035

R:anonymous, Genoscope submitted to the EMBL Data Library, July 1999
 A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.
 A:Reference number: A75001
 A:Accession: G75035
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-331 <KAM>
 A:Cross-references: GB:AJ248287; GB:AL096836; NID:95458657; PIDN:CAB50180.1;
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB1535

G75035 Length: 331 February 11, 2000 15:52 Type: P Check: 3225 ..
 1 MKKVTALIL ASITSLVAL SLGSVRPLS TLNLSLS ISLYRQGLS
 51 GSPYIIIGIR LPRVMALIV GFLALAGTA SOALFRNPPLA DPYIIIGISG
 101 ASIGAAIALA YSPRYEYFA FVGAIMAVYL VYRISKVNH IPVVLIIAG
 151 IAVGFESHAV TSYLLXNRD KVHGLSWLF GTPLATATSK VFIMAVAVGI
 201 GGGMLFMSR ELNLLLGE SIALGDVNL YRLLIIFALA ILTGVAVSES
 251 GIIGFGLVS PHIMRMEYGP NHRRLPVSX MUGGILVLIS DLISRTIVSP
 301 VEIPGVIVTA LFGAPFFAYL LMRKRGELT A

!!AA_SEQUENCE 1.0
 P1:B75110 - lipote-protein ligase A related PAB1916 - *Pyrococcus abyssi* (strain Orsay)
 C:Species: *Pyrococcus abyssi*
 C>Date: 20-Aug-1999 #sequence-revision 20-Aug-1999 #text-change 20-Aug-1999
 C:Accession: B75110
 R:anonymous, Genoscope submitted to the EMBL Data Library, July 1999
 A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.
 A:Reference number: A75001
 A:Accession: B75110
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-249 <KAM>
 A:Cross-references: GB:AJ248285; GB:AL096836; NID:95458067; PIDN:CAB49595.1;
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB1916

B75110 Length: 249 February 11, 2000 15:52 Type: P Check: 5732 ..
 1 MRFPLIVAR PELQMAIDEA ILIARSGKV PPTVALYFK PSSVTIGRFQ
 51 SVRHVDIER AELNIPYVR RITGGGSYFH DEFGITTSV VISDDYHPSL
 101 KNIOESYRFL AGPLVDALKD LGINAFPSGL NDIYVNGKI SGAQTRRRG
 151 ILLQGTIFY STRLEVLAIV LKYSKEKLRD KGISIYERV TILREGIKL
 201 SMOETVELLR NSFRRAPLE EGEITVELE LAQELVEERY GKDEWFMK

!!AA_SEQUENCE 1.0
 P1:E49094 - methylmalonyl-CoA decarboxylase (EC 4.1.1.41) beta chain - *Veillonella parvula*
 C:Species: *Veillonella parvula*
 C>Date: 19-Mar-1997 #sequence-revision 19-Mar-1997 #text-change 26-Feb-1998
 C:Accession: E49094
 R:Huder, J.B.; Dimroth, P.
 J. Biol. Chem. 268, 24564-24571, 1993
 A:Title: Sequence of the sodium ion pump methylmalonyl-CoA decarboxylase from

Veillonella parvula
 A:Reference number: A49094; MUID:94043308
 A:Accession: F49094
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1373 <RND>
 A:Cross-references: GB:L22208; NID:9415592; PID:9415597
 C:Keywords: carbon-carbon lyase; carboxy-lyase
 E49094 Length: 373 February 11, 2000 15:52 Type: P Check: 3704

1 MEAFVAVIOS VINDSGFLAF TTGNAMITLV GILLYLAFRA REEPPLIGP
 51 IAFGLLANT PRNGFEEGYV ALISAGISO IEPPLIFIGV GAMTDFGPI
 101 ANPKTLILGA AAOIGVFAAL GGAMMLGFTA QEAAAGIIG GADGPTIYL
 151 ATRAPHLIG AIAVAASYM SLVPLIOPV MKLETTOKER EIVMEQLREV
 201 TRFKIVEPI VAITFISLL PSITSLGML MGNLFRESG VTRLSDTSQ
 251 NALINTVITF LATGTGLTMS ABHFLSLEI KIIILGLFAR ICGTAGVLF
 301 GKILSLVDGG KTNPLIGSAG VSAVPMARV SOVGAKANP ANFLMHANG
 351 PNVAAGVIGTA VAACTMLAM SNR

11AA_SEQUENCE 1.0
 P1:14222 - NADH dehydrogenase subunit II - Xantusia vigillis mitochondrion
 (Sgc1)
 C:Species: mitochondrion Xantusia vigillis
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14222
 R:Macay, J.R.; Larson, A.; Anandjaya, N.B.; Fang, Z.; Papenfuss, T.J.
 Mol Biol. Evol. 14, 91-104, 1997
 A:Title: Two novel gene orders and the role of light-strand replication in
 rearrangement of the vertebrate mitochondrial genome.
 A:Reference number: 217789
 A:Accession: T14222
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1344 <MAC>
 A:Cross-references: EMBL:U71328; NID:91753276; PID:91753278; PIDN:ABA8290.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Note: ND2
 C:Keywords: mitochondrion
 T14222 Length: 344 February 11, 2000 15:52 Type: P Check: 7879

1 MSPTIQSLIL SSIAIGTITL MTSQSWILAM VOLEINTLAI LPIISKPHF
 51 RATEAMTKYF LQOTAASTMI LFASTLNAMH SGOMINMOLT TEPASTVLT
 101 ALTKKLGLAP VHFMLPEYMQ GSTINTAMIL TWCKLADMT ILINTHONIP
 151 TITLLTIGIT SSIMGCMGL NQOLRKILA YSSIAHLGMI YVAISDETI
 201 TITLLIYIL LSLIFLSMN TSHKIMKDM ATTWLSPTL TTEMULTIS
 251 LGGLPLING LPRWLIQEL TTHHTPIAT LIAFSTILSL FFYRLSYIT
 301 TITLAPNPT MSNKRORCKT KSKSLISTMM FVTTMIPMS PIML

11AA_SEQUENCE 1.0
 P1:S55472 - pectin lyase (EC 4.2.2.10) - Mycosphaerella pinodes
 C:Species: Mycosphaerella pinodes
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Sep-1997
 C:Accession: S55472
 R:Heim, P.
 Submitted to the EMBL Data Library, May 1995
 A:Reference number: S55472

A:Accession: S55472
 A:Molecule type: DNA
 A:Residues: 1299 <HEI>
 A:Cross-references: EMBL:X87580; NID:9854337; PID:9854338
 A:Experimental source: strain Berkeley and Bloxom
 C:Genetics:
 A:Gene: pella
 C:Keywords: 43/3
 C:Keywords: carbon-oxygen lyase
 S55472 Length: 299 February 11, 2000 15:52 Type: P Check: 8743

1 MKEFVSIG LTAITLALP TSMIDVARD VKRAASLED VAIAPERLC
 51 RPSASLPSS RYTRLPPLSS LVLSRALATT SRSAATSPSS DETARSLRT
 101 SPSPSRVSM LSFYLPSPR SSVAALSPSR RYELVTPKEI TYLHRLQSE
 151 NWVDHILDS SPDRDKDY DGLDITHAA DEYVTVNFTL HDHMKASLIG
 201 HSDSNGAEDK GHUTVYANN YLKNUNSRG PSASQATST TTTTTCRMV
 251 STPARVSCIL FRTTILSAPR RLFTRLMDM LLSLATSVM ARTLPRVL

11AA_SEQUENCE 1.0
 P1:D72485 - probable high-affinity branched-chain amino acid transport
 ATP-binding protein APE2524 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: D72485
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no,
 K.; Takahashi, M.; Sekine, M.; Baba, S.; Ankel, A.; Hosoyama, A.;
 Fukui, S.; Nagai, Y.; Nishijima, K.; Yamazaki, H.; Takamida, M.; Masuda, S.;
 Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.;
 Aoki, K.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic
 Crenarchaeon, Aeropyrum pernix K1.
 A:Reference number: A72450; MUID:99310339
 A:Accession: D72485
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1242 <KAW>
 A:Cross-references: DDBJ:AF000064; NID:95105945; PIDN:BA81540.1; PID:dl045326;
 A:Experimental source: strain K1
 A:Gene: APE2524
 C:Genetics:
 D72485 Length: 242 February 11, 2000 15:52 Type: P Check: 7054

1 MEVITILDV RDYKRFSGI EALGVSESV RRGERVGLIG PNGAGKTILF
 51 NIISGIYMPD RGRVYIKGYD ITGMPAYRRS RAGIARTFOI VRPLANTVL
 101 NNVIYGLLR TNDIRPARER AMEALDWGL AGKEDILAND LNLIEKRLE
 151 LARALATPE LLLIDELIAG LRPREVDDLV YTLLEISKRG ITIIMEVHM
 201 RAVMNEAEV IVLHFGKIA EGTPREVASN KLVIEAYMGT GG

11AA_SEQUENCE 1.0
 P1:B71024 - probable lipate protein ligase - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
 C:Accession: B71024
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto,
 S.; Sekine, M.; Baba, S.; Kosugi, H.; Hosoyama, A.; Nagai, Y.; Sakai, M.;
 Ogura, K.; Otsuka, R.; Nakazawa, H.; Takamida, M.; Ohfuku, Y.; Funahashi, T.;
 Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Oguchi, A.; Aoki, K.;
 Yoshizawa, T.; Nakamura, Y.; Robb, F.T.; Horikoshi, K.; Maechi, Y.; Shizuya,
 H.; Kikuchi, H.
 DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3.
A:Reference number: A71000; MUID:98344137
A:Accession: B71024
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-249 <KAW>
A:Cross-references: GB:AP000006; NID:g3236133; PID:d1031537; PID:g3257911
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1487
B71024 Length: 249 February 11, 2000 15:52 Type: P Check: 6905 ..
1 MRPIPLVAR PEVQMAIDEA ILIARSEGRV PDTVRLYFK PSSVTIGRFQ
51 SVAYDNLEK AKELGIPYVR RITGGSVFH DEYGEITYSV VISEDYHPML
101 KNODSYRFL AGPLVDALKD LGIEAFESGL NDIYRGKKI SGNAGTRRRG
151 ITHQGLTMY ATREILGSV LKYSKEKLD KGVKSIPEPV TILEREGIKY
201 GRNEMTEVLR NSFRAFLPD EGELETELE LVEKLIBERY GNDKMFQK
11AA_SEQUENCE 1.0
P1:572288 - ribosomal protein L23 - Plasmodium falciparum plastid
C:Species: Plasmodium falciparum
C:Date: 14-Apr-1998 #sequence_revision 24-Apr-1998 #text_change 07-May-1999
C:Accession: S72288
R:Wilson, R.J.M.; Denny, P.W.; Preiser, P.R.; Rengschardt, K.; Roberts, K.; Roy, J.; Whyte, A.; Strach, M.; Moore, D.D.; Moore, P.W.; Williamson, D.H.
A:Title: Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum.
A:Reference number: S72277; MUID:96346169
A:Accession: S72288
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-75 <MTL>
A:Cross-references: EMBL:X95276; NID:g1171591; PID:e220137; PID:g1171594
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C:Genetics:
A:Gene: rpl23
A:Genome: plastid
A:Note: this apparently degenerate plastid is referred to as the apicoplast
C:Keywords: plastid; protein biosynthesis; ribosome
S72288 Length: 75 February 11, 2000 15:52 Type: P Check: 9584 ..
1 MKVILNPLI YNLFYKINV LNKFCITYSI KFTFLDIKY IINKIRIKI
51 INTNIRIKIN INKNCKLTKY YITFK
11AA_SEQUENCE 1.0
P1:710349 - Very-late factor 1 protein - Oryza pseudotsugata nuclear polyhedrosis virus
C:Species: Oryza pseudotsugata nuclear polyhedrosis virus, OPMNPV
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C:Accession: T10349
R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.
A:Title: The sequence of the Oryza pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.
A:Reference number: Z17011
A:Accession: T10349
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-374 <KAR>
A:Cross-references: EMBL:U75930; NID:g2934903; PID:g1911326

T10349 Length: 374 February 11, 2000 15:52 Type: P Check: 889 ..
1 MDGLGVNNEI TENDWKARIQ SASRFEHVD LATDRQRCIP DEVKNDLSWS
51 KYMPKPEFAP TLKSKYSRL IKIIFSLVEE ADLQNPAYDL DREFDSEYEQ
101 HLAVSPRELQ KMLRELRSVT KETLOLTINE YTNAMNLEPF KIPRWMLPR
151 DKELKIREK EKNFLKNNAI DTILNFDISK ILMNSDYVH DRGLIGAIY
201 FCIMLGTGMR IENARQSLVD DLNVLIKKG LRSDTIGKR KRSRKNLTNN
251 IKTRPLELAR EYARNPTVL QISKNTSTP KDFRRLDEA GVEMERPSN
301 MIRHYLSSNL YNSGVPLQKV ARLNHESPA STKPYLNRYN FDESSDEES
351 GGNNRDSSTG SSANSSSLY QTCG
11AA_SEQUENCE 1.0
P1:578187 - ymf39 protein - Recclinomonas americana (ATCC 50394) mitochondrion
C:Species: mitochondrion Recclinomonas americana
A:Variety: ATCC 50394
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 17-Mar-1999
C:Accession: S78187
R:Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sankoff, D.; Turmel, M.; Gray, M.W.
A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in mitochondria.
A:Reference number: S78127; MUID:97311393
A:Accession: S78187
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-197 <LAN>
A:Cross-references: EMBL:AF007261; NID:g2258325; PID:g2258386
A:Experimental source: ATCC 50394
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
C:Genetics:
A:Gene: ymf39
A:Genome: mitochondrion
C:Keywords: mitochondrion
S78187 Length: 197 February 11, 2000 15:52 Type: P Check: 3067 ..
1 MLSTIQKHV TLTLLVPLVA ILTARDETV HDETLVLLCF ILFLTFLVYA
51 LKDWYASFN DRALQIEKEF NDSYSLEKQT LQLLANHEK QVSLINEIDS
101 LMEFTKSEVN NIQTRQAL RARLISEFT KNTAVKKED AFGFOIQOT
151 NTIYANISLE NISGPTGAL KNTSFEGR ILDEBAISLQ DKEANS
11AA_SEQUENCE 1.0
P1:711851 - ruvB Protein - Thermus aquaticus thermophilus
C:Species: Thermus aquaticus thermophilus
C:Date: 14-Aug-1999 #sequence_revision 14-Aug-1999 #text_change 14-Aug-1999
C:Accession: T11851
R:Tong, J.; Wetmur, J.G.
A:Title: Cloning, sequencing, and expression of ruvB and characterization of RuvB proteins from two distantly related thermophilic eubacteria.
A:Reference number: Z17360
A:Accession: T11851
A:Status: preliminary
A:Residues: 1-324 <TON>
A:Cross-references: EMBL:U22817; NID:g727372; PID:g727373
T11851 Length: 324 February 11, 2000 15:52 Type: P Check: 3584 ..
1 VEDLALPKT LDEYGOERL KOKLRVYLEA AKAREPLEH LLFGPGLG

Mon Feb 14 08:07:32 2000

pir.cat

Page 193

51 KTLIAVIAH ELGVNLRVTS GPAIEKPGDL AAILANSLEE GDILFIDEIH
101 RLSROAEHL YPAMEDEVMD IVIGGPAAR TIRLEPRFA LIGATRPGL
151 ITAPILSRFG IVEHLEYTP BELAQVWRD ARLGVRITE EAALEIGRRS
201 RGTMRVAKRL FRRVRDQAV EGEEVITRR ALEALAAGL DELGLEKDR
251 EILEVILIRF GAGPVGLATL ATALSEDPT LEEVHEPYLI RQGLKRTPR
301 GRVATELAYR HIGYPPVGP ILEP

!!AA-SEQUENCE 1.0
P1:JC5705 - alpha-amylase inhibitor 0.19 - synthetic
C:Species: synthetic
A:Note: Triticum aestivum (common wheat) gene engineered and expressed in
Escherichia coli
C>Date: 03-Sep-1998 #sequence_revision 03-Sep-1998 #text_change 03-Sep-1998
C:Accession: JC5705
R:Okuda, M.; Satoh, T.; Sakurai, N.; Shibuya, K.; Kajii, H.; Samejima, T.
J. Biochem. 122, 918-926, 1997
A:Title: Overexpression in Escherichia coli of chemically synthesized gene for
active 0.19 alpha-amylase inhibitor from wheat kernel.
A:Reference number: JC5705; MUID:98104043
A:Accession: JC5705
A:Molecule type: DNA
A:Residues: 1-128 <OK>
A:Cross-references: DBJ:AB003682; NID:92575815; PID:41020955; PID:92116581
C:Comment: The sequence from wheat (see PIR:A01324) has signal and propeptide
sequences not included in this synthetic sequence.

JC5705 Length: 125 February 11, 2000 15:52 Type: P Check: 3534 ..

1 MSGPMCTPG QAFQVPALPA CRPLRLQCN GSGVPEAVLR DCCQOLAHIS
51 EMCRCGALYS MLDMSYKEHG AOGGAGTGA FPRCRREYVK LTAASITAVC
101 RLPIVVDASG DGAIVCKDVA AYPDA

